An online system for bioterrorism surveillance

by

Melissa Yahya

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Computer Science

Program of Study Committee:
Leslie Miller, Major Professor
Shashi K. Gadia
James A. Roth

Iowa State University

Ames, Iowa

2008

UMI Number: 1450150

TABLE OF CONTENTS

ABSTRACT

Bioterrorism has become one of the greatest threats for nations. As bioterrorism threats are increasing these days, research in bioterrorism surveillance has been conducted to detect the threats at the earliest time. Following this research, some bioterrorism surveillance systems have been developed. Some of these systems use humans as indicators and the others use animals. Both of these systems detect the threats by collecting and analyzing the statistical data on health trends. Since most bio-weapons start as animal diseases, using animals as indicators allow identification of a bioterrorism threat at earlier time than using humans.

Current bioterrorism surveillance systems based on animals focus only on certain groups of animal species while there are other animal species that are also very susceptible to the highly-threatening diseases. The system introduced here provides a step towards the identification of highly-threatening diseases in animal species, such as cattle and pigs. In the event of such identification, a red flag is raised and related information is sent to agents responsible for looking into the threats in more detail. The design, implementation and result of this system are given and explained.

1. INTRODUCTION

As terrorism threats are increasing, there is an urgent need to be able to identify any possible terrorism act in a timely fashion to prevent it from happening, or if that is impossible, reduce the losses. The greatest threat among terrorism actions comes from bioterrorism. Using biological agents, these bio-weapons will disperse quickly and eventually they have the potential to cause a nationwide epidemic while terrorism actions such as suicide bombings using conventional explosives or chemical weapons happen at some particular area and will stop at ground zero [4]. Based on the fact that most of the diseases used as bio-weapons are animal diseases, these diseases will be seen in animals first. In addition, bio-weapons may be used to destabilize the food system as a mean of causing panic. Hence, we will focus on using animal diseases to identify possible bioterrorism threats.

Information is very important in our current world. With the rapid growth of available information in every aspect of our life, every crucial decision made relies on the available information. It is no wonder that so many search engines have been developed to support these needs (e.g., Google, Ask.com, Yahoo! Search). Using these search engines, people can look up information by specifying the topic of their interest and the search engines will give a list of websites that have the relevant information.

Most of the existing bioterrorism surveillance systems detect the occurrence of bioterrorism by collecting and analyzing the statistical data on health trends. For systems that use humans as indicators, this data is collected from hospitals. Similarly in the case of animals, data from pet hospitals is collected. Since most bio-weapons are likely to be animal diseases, using animals as indicators is a good method of early identification of any possible bioterrorism threats. The problem with using animals as indicators is the fact that not all sick animals are brought to pet hospitals by their owners. Instead, some of them try to treat their animals by themselves, using a search engine's ability to easily find information online.

Here, we have developed an online system that uses information given by the user to try to detect whether a bioterrorism attack has occurred. When an animal owner or a veterinary is looking for some information about symptoms his/her animal has, the tools at our website will try to provide them with information about the disease(s) that match the symptoms entered by the user. In addition, the system attempts to determine whether these symptoms are related with some highly threatening diseases.

The tools at our website will let the user input the species of his/her pet or livestock and choose the symptoms from a list. Based on the input given, the user will receive a set of relevant documents. In addition, using the symptoms provided, the system will try to identify whether the symptoms are related to diseases that are considered possible bioterrorist threats, such as anthrax, botulism, brucellosis, plague, q-fever, tularemia [1]. If the system determines that the selected symptoms indicate any of these diseases then it will raise a red flag to notify the appropriate individuals. Even though it may not be caused by a bioterrorism action, some procedures may still be needed to prevent the spread of a highly contagious disease.

The next chapter looks at a brief review of the relevant literature. Chapter 3 provides an overview of the system model. Chapter 4 provides an in depth review of our approach to do the red flag analysis. We briefly look at some of the relevant implementation issues in Chapter 5. Conclusions and future work are addressed in Chapter 6.

2. BACKGROUND

A great deal of research has been conducted into ways of dealing with bioterrorism so that it is possible to identify threats at the earliest time. Instead of waiting for a confirmed diagnosis as is generally done in traditional disease surveillance systems, researchers and public health officials have developed a method called *syndromic surveillance* which involves collecting and analyzing statistical data on health trends [15].

One of the examples of a *syndromic surveillance* system is the Comprehensive Assessment for Tracking Community Health (CATCH) data warehouse with real-time flash components which currently are implemented in Florida [8]. The system uses the historical data that hospitals have and uses the patterns found in these data to determine whether or not they need to raise an alarm. This is done by checking whether the number of occurrences of a disease is drastically greater than expected over a certain unit of time, then an alarm will be raised. Here, the system identification for bioterrorism relies on the number of the patients that come to the hospitals with some particular symptoms or diseases.

Most of the bioterrorism surveillance systems that have been developed to date have been focused on humans, since human health has always been our priority. It is also a good indicator since a person will go to the doctor when he/she is not feeling well and the hospital will have relatively complete medical records for this person. As a result, it will be easier to identify when some symptoms are unusual and may be related to bioterrorism. However, it might be a little too late by the time that they have identified the case because of the fact that this person may have been interacting with other people and traveling. As a result, the disease may have been spread over a significant portion of the population.

Based on the fact that most bio-weapons are likely to be animal diseases; such as anthrax, avian influenza, and SARS; some scientists are considering using animals for early detection of bioterrorism attacks. In addition, bio-weapons may be used to destabilize the food system as a means of causing panic. For some bioterrorism agents, animals are more susceptible than humans because the incubation period of certain agents in animals is shorter than in humans, animals have a higher possibility of being exposed at a more intense level than humans, and animals are more susceptible [13]. Hence, there have been some systems developed that focus on animal diseases. One of them is the Veterinary Medical Database-Surveillance of Syndromes (VMD-SOS) developed at Purdue University [10]. It identifies bioterrorism threats using abnormal laboratory findings and the clinical signs derived from a pet hospital, in order to identify any serious disease and whether the occurrences of this disease exceed the normal frequency in a fixed period. VMD-SOS system basically focuses on pets, with the assumption that some diseases such as tularemia and anthrax can be transmitted from the pet to the owner. Their work is interesting because there is a large population of domestic pets in the United States. Using the fact that more than 75% of these pets are brought to the veterinary routinely, VMD-SOS uses the data from the pet hospital to get early warning of possible bioterrorism attacks.

The CDC publication Biological and Chemical Terrorism: Strategic Plan for Preparedness and Response gives a list of the biological agents that have a high possibility of being used in bioterrorism attacks [3]. They categorize these critical biological agents based on risk into categories A, B and C. Category A consists of natural agents which are easily disseminated, have high mortality rate, and can cause public panic. The agents in this category are:

- variola major (smallpox)
- *bacillus anthracis* (anthrax)
- *yersinia pestis* (plague)
- *clostridium botulinum* toxin (botulism)
- *francisella tularensis* (tularemia)
- filoviruses
- arenaviruses

Category B also consists of natural agents that are moderately easy to disseminate, have moderate morbidity and low mortality rate. Category C consists emerging pathogens that can be disseminated in the future because of factors such as availability, ease of production and dissemination, and potential for causing high morbidity and mortality rate and major health impact.

Rabinowitz et. al. found that animals such as sheep, horses, cattle, cats can provide early warning of bioterrorism attacks [13]. They found that sheep and cattle can give early warning of acute bioterrorism attack using anthrax, while dogs and pigs are not inherently good indicators. However, certain behavioral characteristics of these animals (for example, dogs playing in the contaminated mud) tend to make them more susceptible to pick up anthrax through minor cuts and bruises; thereby making them good indicators [5]. They have also suggested that early warning for the plague, can be detected from cats. For category B agents, early warning of an acute bioterrorism attack in the form of a foodborne illness is possible to detect in cattle. In addition, alphaviruses can be detected in horses. Rift valley fever can be detected in cattle and sheep. For category C agents, wild birds can be used for early warning of the presence of the flavivirus.

The difficulty of using animals for early warning of a bioterrorism attack is that, they might not be brought to an animal hospital when they are sick. Sometimes the owner may only call a veterinary to check on his/her animals or worst by doing nothing until it is too late. In the VMD-SOS system, the focus is limited to pets, especially dogs and cats, while other animals such as horses, cattle and sheep are not considered. However, these animals are especially important for attacks that are designed to destabilize the food supply.

Los Angeles County Public Health's (Veterinary Public Health) Animal Disease Surveillance System (http://www.lapublichealth.org/vet/disintro.htm) provides users (non-experts and veterinaries) with the ability to report diseases of animals to the Los Angeles County Department of Health Services Public Health Veterinary Unit. Their system also provides information about some high-threat disease outbreaks. The problem with this approach is that in many cases animal owners do not realize that the disease their animals have is a high-threat; hence they may not feel the necessity to submit such report. Animal owners may also fear government reaction to such a report.

In this project, we are building a system where the user can search for literature to get some information about the animal diseases. Most of the time, when the user has no clue about the disease that his/her animal suffer, he/she will try to search for some information based on the animal's symptoms before deciding whether there is a need to bring the animal to the hospital. Sometimes he/she will try to give simple and inexpensive treatments rather than bring it to the hospital, in order to save money. Using this system, the user specifies the symptoms of the animal and its species, to get some information on what might be wrong with the animal. If the symptoms are identified as a high-threat disease, the system will automatically raise a red flag to notify the persons in charge so they can initiate the needed actions using the contact information of the user. This way whenever any high-threat disease is identified, it can be prevented from spreading by a quick response.

For an illustration of how our system works, let's say that there is a user Tom Johnson that has some cattle. Today, one of his cattle is found dead with bloody discharges from its nose, mouth and anus. He noticed that yesterday his cattle seemed excited and then depressed, disoriented and short of breath. He tries to get some information about his cattle using a search engine in the internet which leads to our system. He logs in into our system and inputs his query by selecting "cattle" for species and "excitement followed by depression", "disoriented", "shortness of breath" for part of symptoms when the animal is alive, "found death", "bloody discharges from orifices" for part of symptoms when the animal is dead. If he knows of any persons that have been closely associated with the animal that show any signs or illness, he can also select the appropriate symptoms from the list for humans.

Our system returns to the user a set of documents that have information related to the symptoms that have been entered and the designated species while also processing the information to identify that anthrax, a highly-threatening disease, may have occurred. The system raises a red flag and notifies the person in charge by sending email about the contact information of Tom Johnson, the disease that has been identified (in this case anthrax), the most relevant document that has been given to Johnson, and information about given symptoms and species. This way, the individual responsible for investigating the red flag can contact Johnson to follow up about his cattle and determine whether further actions are needed.

In the next chapter, we examine the model of our complete system.

3. SYSTEM MODEL

In the system proposed in this paper, using the web browser, the user inputs the species and the symptoms of his/her pet or livestock and chooses the relevant symptoms that his/her animal has. The system will search based on the symptoms for relevant documents from data sources and return them to the user. To determine which data sources to use, the system first checks each data source to see if the user has the required permission to access it. It also checks whether the data source contains information about the species given from the data source's description. In this chapter, the architecture of this system will be given, followed by descriptions about the components of the system and their interaction as a whole system. A block diagram of the proposed system is shown in Figure 3.1.

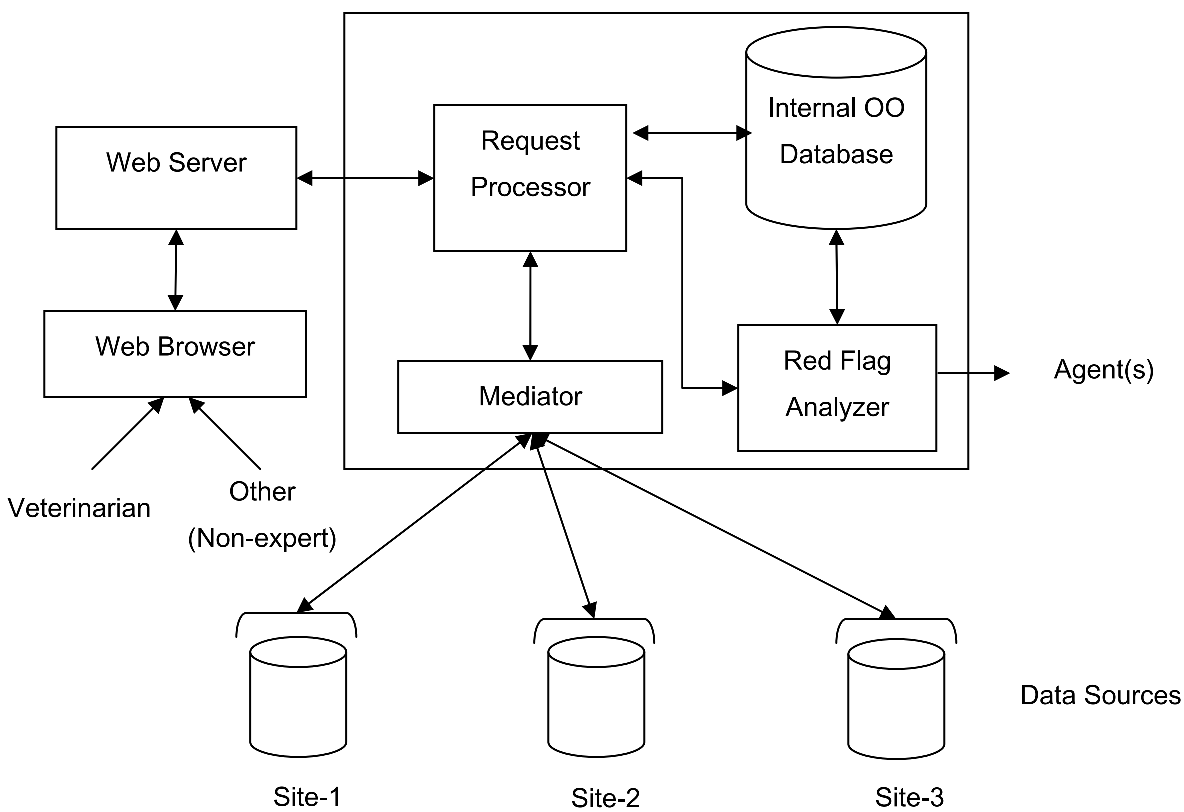

Figure 3.1: Model of the system architecture.

3.1 Web Browser

The *web browser* is the part of the system that the user will directly interact with. Using his/her *web browser*, after login, the user can update their contact information, search for relevant information by providing the information about his/her animal species and the symptoms that his/her animal suffers and receive a set of documents describing the possible diseases and treatments.

3.2 Request Processor

This component is responsible for processing a user's request. After a user submits a query from the *web browser*, the system creates a request object, which is an instance of class **RequestDocument**. The *request processor* takes this object and initiates the appropriate action by first checking the user's credentials. Additionally, based on the species that the user specifies in the input, the *internal object oriented database*, which contains a description of the data sources, is queried to decide which data sources are pertinent. Next, the relevant documents have to be retrieved from the data sources. The request object notifies the *mediator* to dispatch the request to the selected data sources. At the same time, it initiates the *red flag analyzer* process to look into the possibility of the occurrences of a highly-threatening disease from the symptoms given. Once the *mediator* receives all the results from data sources, the request object reads them from the *mediator*. Finally, these documents are returned to the user through the *web browser*.

3.3 Mediator

Our *mediator* has the responsibility of mediating the interaction between the *request processor* and the data sources. Once a *mediator* receives a request from the *request processor*, it dispatches the request to the selected data sources. It sends the request to the *wrapper* of each selected data source. Then, this *wrapper* calls the corresponding methods in each site using the data source's *local interface view (LIV)*. This *LIV* allows the *wrapper* to call methods from remote site as if it is local. The structure of a data source with its *wrapper* and *LIV* is shown in Figure 3.4.

Once the data source's *wrapper* obtains the result, it will send it to the *mediator*. The *mediator* gathers all the results returned from those data sources, and returns it to the *request processor*. The *request processor* will show the results to the user through the *web browser*. The *mediator* acts as the glue that integrates all data sources to our system.

3.4 Red Flag Analyzer

The *red flag analyzer* is the part of the system that is responsible for analyzing the likelihood of the occurrence of a highly-threatening disease. For each highly-threatening disease, the *red flag analyzer* calculates the similarity between its symptoms and the symptoms given by the user. If there are sufficiently high similarities between those symptoms then there is a high possibility that this disease may be the reason for the animal's illness and some further actions are needed as soon as possible. The *red flag analyzer* will raise a red flag; notify the person(s) in charge that such disease has been identified by sending email to him/her with the information about the user, the disease, and the most relevant document. Hence, he/she can investigate the likelihood of the disease being the problem. The user can be contacted to gain more information if necessary. Further details about *red flag analyzer* will be given in Chapter 4.

3.5 Internal Object Oriented Database

The *internal object oriented database* contains information about users of our system, data sources integrated with our system, universities that own those data sources, diseases that are considered to be highly threatening, animal species that can be infected by those diseases, and symptoms of animal diseases where some of these symptoms can be an indication that a highly-threatening disease occurs. The types of information stored in the *internal object oriented database* can be seen in Figure 3.2. The *internal object oriented database* is composed of an object-relational mapping tool (ORM) at the front end, and a relational database at the backend. The relational database physically stores the data used by the system, while the ORM tool is used to represent and handle data as objects. Relational databases will be described in more detail next, while Hibernate, ORM tool used, will be discussed in Chapter 5.

3.5.1 Relational Database

A relational database stores objects (*entities*) with their properties (*attributes*) and the relationships between these objects. These objects are represented as tables; some examples from Figure 3.2 are *User*, *University, DataSource, Symptoms*, and *Diseases*. They may relate one to another through common values of their attributes.

- A primary key (PK) is a key that uniquely identifies each record in a table. It may consist of one attribute or more.
- A foreign key (FK) is an attribute (or attributes) of a table that links to the primary key of another table. Thus, it creates the relationship between tables.
- A relationship is a link between two tables which make it possible to find data in one table that pertains to a specific record in another table. Based on the cardinality, there are 3 types of relationships; they are one-to-many, one-to-one, and many-to-many. In Figure 3.3, we can see an example of a one-to-many relationship between the tables *University* and *DataSource*. It means that one *University* can have more than one *DataSource* and a *DataSource* is owned by at most one university. The *dash* symbol used in the relationship represents "one", and the *ring* symbol represents "zero". *Dash* and *dash* represents exactly one. A *crow's foot* in the relationship represents "many". Using these symbols, we can see from Figure 3.3 that a *University* may have zero or more *DataSource* and a *DataSource* is owned by exactly one *University.*

3.5.2 User Information

We store all the information about the users of our system in table *User*. This table has the attributes *user_id, name, username, password, user_group, email, address, city, zip, phone,* and *univ_id.* The primary key of *User* table is *user_id* and the foreign key is *univ_id.*
When a red flag is raised, the user information such as name, email, address, city, zip code, and phone number are sent to the person in charge. Using this information, he/she can contact the user to get more information about his/her animal's illness. Information such as username, password and group of the user is used by the system to authenticate a user.

3.5.3 Data Source Information

The *DataSource* table stores information about data sources integrated with our system. This table has attributes *datasource_id* (primary key), *URL*, *name,* and *univ_id* (foreign key). The foreign key *univ_id* allows the system to check the university that owns the data source. Only users from the university (i.e., they have the same value of *univ_id*) are allowed to access this data source.

To describe the contents of each of these data sources, we use the *DataSource_Description* table. It stores information about which animal's species are described by the documents in a data source. Our system uses this information to decide where to find documents relevant for users.

3.5.4 Disease Information

Table *Disease_Species_Symptom* has the attributes *disease_id, species_id, symptom_id, weight, sign, and status*. Its primary key is formed from *disease_id, species_id* and *symptom_id*. These attributes are also a foreign key of this table. Table *Disease_Species_Symptom* contains information about contribution of a symptom on a certain species towards the occurrence of a highly-threatening disease. The portion of the contribution is expressed by an attribute *weight* and the category of the contribution is expressed by the attribute *sign*. The attribute *status* shows whether a symptom of a disease in a certain species of animal appears when the animal is alive or when the animal is dead. The value of attribute *status* of a symptom is 0 if the symptom is noticeable when the animal is alive. If the symptom is noticeable after the animal is dead then the value of attribute *status* is 1. If the presence of a symptom emphasizes the occurrence of a disease then the value of attribute *sign* of this symptom for this disease is 1. If the presence of a symptom emphasizes that a disease does not occur then the value of attribute *sign* of this symptom for this disease is 2.

The information of human symptoms is also stored in table *Disease_Species_Symptom*. For humans, the value of attribute *species_id* = 1 and *status* = 2.

Table Diseases

disease_id	Name
1	Anthrax
2	Botulism
3	Plague
4	Tularemia
5	Brucellosis

Table Species

Species_id	name
1	human
2	cattle
3	sheep
4	goats
5	Pigs

Table Symptoms

symptom_id	Name
1	Found dead
2	Fever
3	Shivering trembling
4	Shortness of breath
5	Congested mucous membranes
6	Bloody discharge from nose mouth and anus
7	Swellings on the body (e.g. ventral neck thorax shoulders)
8	Excitement followed by depression
9	Stupor
10	Disorientation
11	Abortion
12	Milk production drop severely
13	Singeing of the hair (by a history of electrical storm)
14	Crepitating swelling
15	Reduced fertility
16	Weakness in young

Table Disease_Species_Symptom

disease_id	species_id	Symptom_id	weight	sign	Status
1	2	1	0.8	1	1
1	2	2	0.2	1	0
1	2	3	0.3	1	0
1	2	4	0.3	1	0
1	2	5	0.3	1	0
1	2	6	1	1	1
1	2	13	1	2	1
1	2	14	1	2	0
1	2	14	1	2	1

Figure 3.2: Table of diseases, species, symptoms and their relationships.

As an example, the fifth row of table *Disease_Species_Symptom* in Figure 3.2 describes a symptom “Congested mucous membranes” of “anthrax” in the animal species “cattle”. This symptom is noticeable when the animal is alive and the occurrence of this symptom suggests the possibility that “anthrax” has occurred. Compared to the other symptom “Fever”, based on the value of the attribute *weight*, the occurrence of this symptom gives slightly higher belief that the disease that may have occurred is “anthrax”.

The *red flag analyzer* uses the information stored in table *Disease_Species_Symptoms* to analyze whether or not a highly threatening disease has occurred. *Request processor* uses the information about data sources to decide which data sources that a certain user has access to. Data source information is also used in deciding which data sources are relevant.

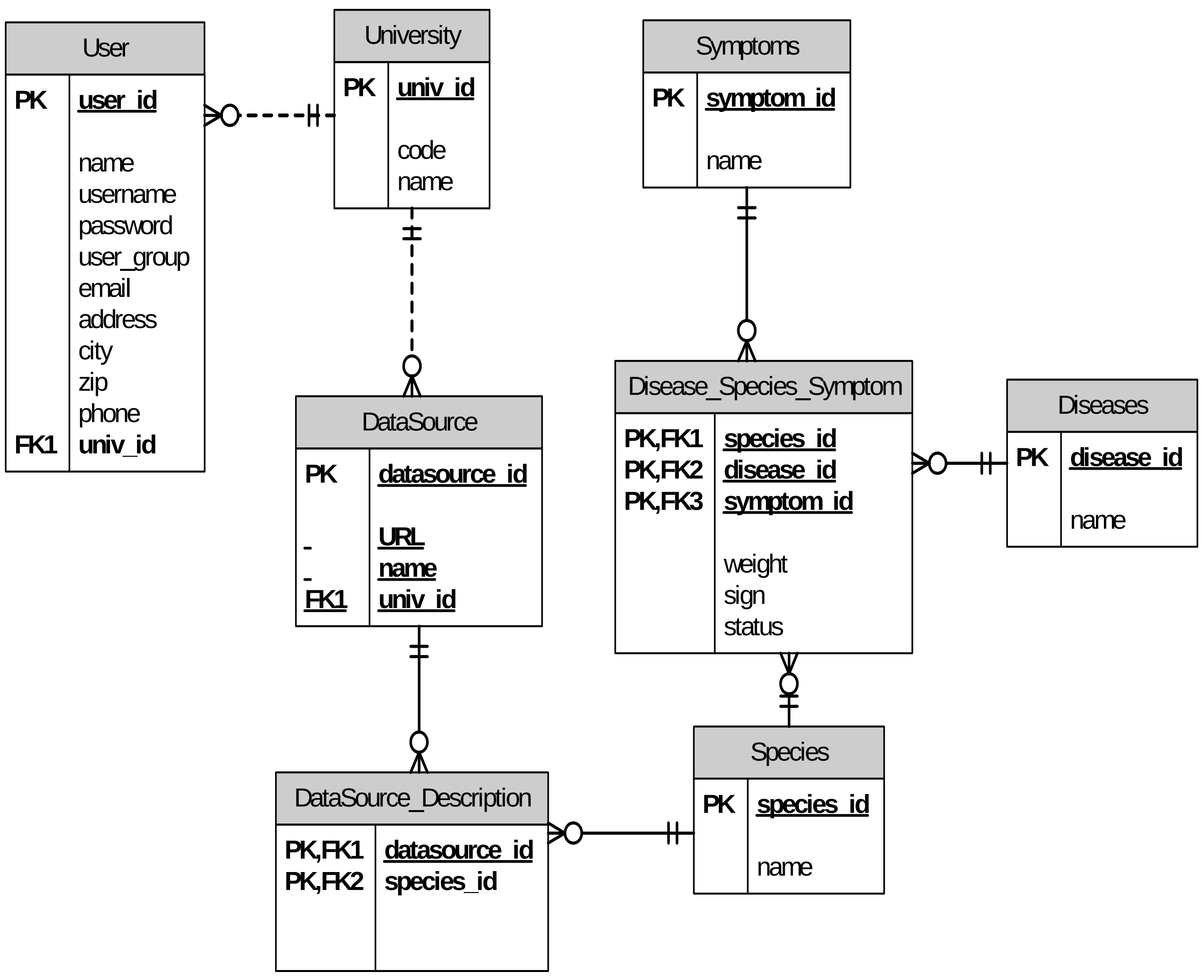

Figure 3.3: Design of *Internal Object Oriented Database* back end.

3.6 External Data Sources

The external *data sources* contain information about animal diseases. For simplicity, we assume each of them is located on different site. In each site, there are two functionalities available: indexing and matching. Indexing is a process to retrieve the important keywords from documents and matching is to calculate the similarity between these keywords to a query given by a user. The indexing process is done everyday to make sure that the indices are always up to date and the information is saved in a file. The matching function uses this index information whenever there is a request for information. Anytime there is a request for information from the *mediator*, the matching function is executed and then from the results, a set of documents containing the required information will be returned.

We can see the structure of an external data source from Figure 3.4. A *mediator* sends a user query to a data source's *wrapper*. Then this *wrapper* continues by sending the request to *Data Source LIV*. This *Data Source LIV* allows our system to use the data source's local tools. Once the *wrapper* receives the result, it sends it back to the *mediator*.

The next chapter will gives further details about the *red flag analyzer*.

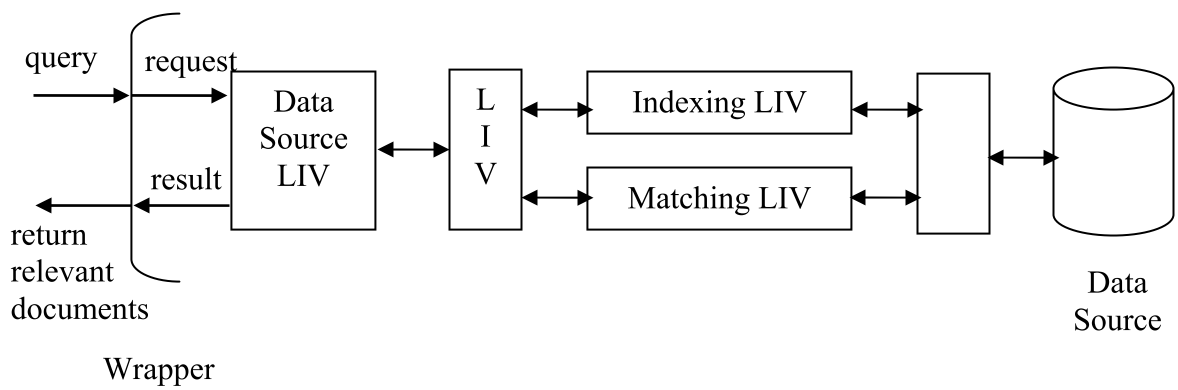

Figure 3.4: *External Data Source* layout.

4. RED FLAG ANALYZER

In this chapter, we will look into *red flag analyzer* in more detail. As mentioned in the previous chapter, this component is responsible for analyzing the likelihood of a highly-threatening disease occurrence. There are two possible approaches; the first one is using a data mining approach to look for patterns from a large amount of data about sets of symptoms related to diseases in some species of animal and the second is by analyzing a species of animal one instance at a time to determine whether the current set of symptoms constitutes an instance of a certain disease.

In the data mining approach, the ability to identify the occurrence of a disease based on the symptoms pattern really depends on the training data and the contribution of a symptom to the disease identification itself is not that obvious. This approach needs a large amount of training data that should be representative of a wide range of diseases. The main problem for this approach is the fact that some highly threatening diseases do not occur that often. As a result it is impossible to get a large amount of training data that is sufficient to cover all of the highly-threatening diseases that need to be identified using the data mining approach. In addition, when a new highly-threatening disease emerges, it is impossible to find a large amount of training data that covers this disease sufficiently in a short time while we want to be able to react quickly to the occurrences of such diseases.

In our system, we use the second approach where for each species of animal, we analyze whether the symptoms given by a user are sufficiently similar with the set of symptoms related to a particular disease. Using this approach, for each highly-threatening disease we want to identify, we store the information needed about this disease in the *internal object oriented database*. When a new highly-threatening disease emerges, we can add information about this disease in the *internal object oriented database* as soon as it is available.

Using the symptoms selected by the user that describes his/her animal symptoms and the species that he/she has, the *red flag analyzer* of our system calculates the similarity of these symptoms against a set of symptoms for each highly-threatening disease for the species that he/she has named. If there is a sufficiently high similarity then a red flag email contains information of the user, the disease and the symptoms that trigger the red flag will be sent to the person in charge to notify him/her that a highly-threatening disease may have occurred. Using our approach, we try to give all of the relevant information to the person in charge to help him/her re-analyze the reason that a red flag is raised if he/she has to do so.

As mentioned before, for each disease that is considered to be highly threatening, we store a set of symptoms related to this disease for each species of animal in the *internal object oriented database*. We classify symptoms in this set into 2 categories, symptoms when the animal is alive and symptoms for the case when the animal is dead. As example, in cattle that have anthrax, symptoms such as shivering, trembling, excitement followed by depression, shortness of breath can be seen when it is alive, while symptoms such as bloody discharge from the body openings can only be seen upon its death. In some cases, such as anthrax, the owner typically doesn’t notice that the animal is sick until it is found dead. The reason is there is only a short period of time between the time when first symptom appears and the time when it is found dead. Hence, we calculate the red flag value of a set of symptoms from the first type separately from the second. The highest red flag value is used to decide whether a red flag needs to be raised.

Based on the contribution type of a symptom to a disease, there are 2 groups of symptoms; the first one is a group of symptoms that indicates a certain type of disease may have occurred by its presence, the second is a set of symptoms that indicates a predicted type of disease does not occur by its presence. In cattle, symptoms such as shivering, trembling, excitement followed by depression, shortness of breath belong to the first group of anthrax’s symptoms while a symptom such as crepitating swellings belongs to the second group. By finding crepitating swellings on cattle, along with other symptoms similar to those given above, we can conclude that the disease is likely to be blackleg, not anthrax, because crepitating swellings does not occur in anthrax.

Some diseases are transmitted from sick animals to humans through direct contact. When it is known that there are any humans that have been in contact with a sick animal and showing symptoms of disease similar to the one predicted for the animal, it increases the suspicion that the predicted disease has occurred. However, such information is not always available and not all diseases have the same means of transmission. Hence, we can’t always use this information in every red flag value calculation. When the given symptoms of animal are not sufficient to predict whether a highly-threatening disease occurs, we use human symptoms (when available) as well to check whether a red flag need to be raised.

4.1 Method

To calculate a red flag value for a disease in a species of animal, we use the formulation given below as the matching function:

$$red\_flag = \frac{\sum_{i=1}^{n} (w_i * s_i)}{\sum_{j=1}^{n} (w_j * d_j)} - \sum_{k=1}^{m} (w'_k * s'_k)$$

In the formulation above to calculate a red flag value of disease dis_x in species of animal $spec_y$, w_i is a weight that indicates the importance of symptom sym_i in identifying disease dis_x in species of animal $spec_y$, w'_i is a weight that indicates how big the influence of the presence of symptom sym'_i in invalidating the predicted disease dis_x, n is the total number of symptoms in the first group to identify the occurrence of disease dis_x in species of animal $spec_y$, and m is the total number of symptoms in the second group to invalidate the prediction that disease dis_x has occurred in species of animal $spec_y$. For each symptom sym_i of disease dis_x in species of animal $spec_y$, the value of d_i is 1 if sym_i belongs to the first group of dis_x's symptoms in $spec_y$, that is a group of symptoms emphasize a disease may have occurred by their presence, otherwise d_i is 0. If this symptom sym_i is not only in the first group, but also in the list of user symptoms then the value of s_i is 1, otherwise it is 0. If symptom sym'_i of disease dis_x in species of animal $spec_y$ belongs to the second group and selected by the user, a group of symptoms imply that a predicted disease does not occur by their presence, then the value of s'_i is 1, otherwise it is 0. Note that the first expression uses unnormalized weights ($w_i + w_j$), where as the second component uses normalized weights (w'_k).

The weights of symptoms in the first group are assigned based on how important and unique the symptoms for identifying a disease in a species of animal. The more unique a symptom for a disease, the higher the value we need to assign for its weight. Another factor is the fact that some symptoms are more likely to be noticed than others. Hence, there is a higher chance that these symptoms are mentioned by the user than the other. The weights for these symptoms should be higher than for the symptoms that are hard to notice.

The weights for symptoms in the second group are assigned based on how big the influence of their presence is in invalidating the predicted disease. While symptoms in the first group together constitute to the occurrence of a particular disease, the symptoms in the second group might represent the occurrence of different diseases. As example, “Singeing of the hair (by a history of electrical storm)” and “Crepitating swelling” (Figure 4.1) are not representative of the same diseases and they are caused by different reasons. The first one happens to cattle when they are struck by lightening in an electrical storm, while the second is because of blackleg. Hence, these symptoms most likely won’t appear together.

When a symptom in the second group of symptoms is the only symptom that suggests the occurrence of disease dis_y, a disease that is different than the predicted one dis_x, then we can assign its weight with some value where 1 is the maximum. If there are more than one symptom that suggests the occurrence of a different disease dis_z (the predicted disease is dis_x), the total of these symptoms’ weights should be at most 1. The assignment of these symptoms’ weights (symptoms of disease dis_z) depends on how importance each of these symptoms is compared to the other symptoms of disease dis_z. However, we usually see the first case where each symptom in the second group invalidates our predicted disease by referring to different diseases.

In Figure 4.1, we can see the assignment of weights for anthrax’s symptoms in cattle. We give a weight of 0.8 for the symptom “Found dead” and a weight of 1 for “Bloody discharge from nose, mouth and anus”. Here, we assign a higher value for the weight of “Bloody discharge from nose, mouth and anus” than the weight of “Found dead” because it gives more information to differentiate which disease has occurred. For symptoms such as “Fever”, “Abortion”, and “Milk production drop severely”, we assign a very low value for the weight because these symptoms provide little information that anthrax has occurred. Their weights are the lowest because they are not as unique as the other symptoms. We can find these symptoms in some other diseases. We give the weights of symptoms “Swellings on the body (e.g. ventral neck, thorax, shoulders)” and “Excitement followed by depression” a high value because they are very unique and noticeable compared to the other symptoms.

Compared to others, their weights are the highest because they give more credence to the fact that anthrax has occurred.

No	Id	Symptom	Weight	Group	Stat
1	1	Found dead	0.8	1	1
2	2	Fever	0.2	1	0
3	3	Shivering, trembling	0.3	1	0
4	4	Shortness of breath	0.3	1	0
5	5	Congested mucous membranes	0.3	1	0
6	6	Bloody discharge from nose, mouth and anus	1	1	1
7	7	Swellings on the body (e.g. ventral neck, thorax, shoulders)	0.4	1	0
8	8	Excitement followed by depression	0.4	1	0
9	9	Stupor	0.3	1	0
10	10	Disorientation	0.3	1	0
11	11	Abortion	0.2	1	0
12	12	Milk production drop severely	0.2	1	0
13	13	Singeing of the hair (by a history of electrical storm)	1	2	1
14	14	Crepitating swelling	1	2	0
15	14	Crepitating swelling	1	2	1

Figure 4.1: Table of anthrax's symptoms in cattle.

For anthrax's symptoms in the second group, we assign value 1 for the weights of "Singeing of the hair (by a history of electrical storm)" and "Crepitating swelling". The reason is the occurrence of any of these symptoms invalidates the speculation that anthrax has occurred. We can assign the value 1 for weight of these symptoms because they are caused by different reasons. The first symptom is caused by a lightning, while the second is caused by the *Clostridium chauvoei* bacteria and indicates that the blackleg disease may have occurred. To explain how to calculate a red flag value of a disease, we will use the example below.

Example:

A user uses our system and selected the following symptoms.

Symptoms of cattle when it is alive:

- shivering, trembling
- shortness of breath

- excitement followed by depression
- disorientation

Symptoms of humans that have direct contact with the cattle:

- raised itchy bump develops into a vesicle
- a painless ulcer with a black center

Using the formulation above, our system will calculate the red flag value of each highly-threatening disease for these symptoms. For this example, let's say now we are calculating the red flag value of these symptoms for anthrax. Since the user didn't specify symptoms of the cattle when it is dead, we only calculate the red flag value for symptoms of the cattle when it is alive.

The red flag formulation is:

$$red\_flag = \frac{\sum_{i=1}^{n} (w_i * s_i)}{\sum_{j=1}^{n} (w_j * d_j)} - \sum_{k=1}^{m} (w'_k * s'_k)$$

The weights for given symptoms from the table in Figure 4.1:

Symptoms of animal:

- shivering, trembling : 0.3
- shortness of breath : 0.3
- excitement followed by depression : 0.4
- disorientation : 0.3

Symptoms of humans:

- raised itchy bump develops into a vesicle : 0.3
- a painless ulcer with a black center : 1.5

We calculate the red flag value of anthrax as following:

$$W_1 = \frac{\sum_{i=1}^{n} (w_i * s_i)}{\sum_{j=1}^{n} (w_j * d_j)}$$

$$= \frac{(0.2*0)+(0.3*1)+(0.3*1)+(0.3*0)+(0.4*0)+(0.4*1)+(0.3*0)+(0.3*1)+(0.2*0)+(0.2*0)}{(0.2*1)+(0.3*1)+(0.3*1)+(0.3*1)+(0.4*1)+(0.4*1)+(0.3*1)+(0.3*1)+(0.2*1)+(0.2*1)}$$

$$= \frac{1.3}{2.9}$$

$$= 0.448275862$$

$$W_2 = \sum_{k=1}^{m} (w'_k * s'_k)$$

$$= (1*0)$$

$$= 0$$

red_flag = $W_1 - W_2 = 0.448275862$

Using the cattle's symptoms given by the user, we get the red flag value of anthrax as 0.448275862.

Having this result, we check whether this value is greater or equal to a threshold of 0.45 as we have found that this threshold gives the best results for our limited data in identifying the highly-threatening diseases. If it is, an email contains information about the user, symptoms, suspected disease and the most relevant document are sent to the person(s) in charge. Otherwise, if the human symptoms are given then we recalculate the red flag value using the formulation as following:

$$red\_flag = \left(\frac{\sum_{i=1}^{n} (w_i * s_i)}{\sum_{j=1}^{n} (w_j * d_j)} - \sum_{k=1}^{m} (w'_k * s'_k) \right) + \left(\frac{\sum_{i=1}^{p} (wh_i * sh_i)}{\sum_{j=1}^{p} (wh_j * dh_j)} - \sum_{k=1}^{q} (wh'_k * sh'_k) \right)$$

The first part of this formula is the same as before; wh_i is a weight that indicates the importance of human symptom sym_i to disease dis_x, wh'_i is a weight that indicates how big the influence of the presence of human symptom sym'_i in invalidating the predicted disease dis_x, p is the total number of human symptoms in the first group to identify the occurrence of disease dis_x, q is the total number of human symptoms in the second group to invalidate the prediction that disease dis_x has occurred. For each symptom sym_i, the value of dh_i is 1 if it belongs to the first group, otherwise it is 0. If symptom sym_i is not only in the first group, but also in the list of user symptoms then the value of sh_i is 1, otherwise it is 0. If symptom sym_i

belongs to the second group and selected by the user then sh'_i is 1, otherwise it is 0. Note that the first expression uses unnormalized weights ($w_i + w_j + wh_i + wh_j$), where as the second component uses normalized weights ($w'_k + wh'_k$).

As we can see from Figure 4.2, symptom “A painless ulcer with a black center” has the highest value for its weight. The occurrence of this symptom is a strong indication that anthrax may have occurred. On the other hands, the symptom “Flu-like signs” has the least weight value. The reason is this symptom doesn’t give a lot of information about the disease. It occurs in many human diseases such as smallpox, chicken pox, tularemia, brucellosis, food-borne illness, etc.

No	Id	Symptom	Weight	Group	Stat
1	35	Raised itchy bump develops into a vesicle	0.3	1	2
2	36	A painless ulcer with a black center	1.5	1	2
3	37	Swelling lymph glands in the adjacent area	0.7	1	2
4	38	Vomiting of blood	0.7	1	2
5	39	Diarrhea	0.3	1	2
6	40	Severe breathing problems	0.7	1	2
7	41	Shock	0.6	1	2
8	42	Flu-like signs	0.3	1	2

Figure 4.2: Table of anthrax’s symptoms in human.

Using our example before, we recalculate the red flag value of anthrax using weights from table given in Figure 4.1 for the cattle symptoms and weights from table given in Figure 4.2 for the human symptoms as following:

$$red\_flag = \left(\frac{\sum_{i=1}^{n} (w_i * s_i)}{\sum_{j=1}^{n} (w_j * d_j)} - \sum_{k=1}^{m} (w'_k * s'_k) \right) + \left(\frac{\sum_{i=1}^{p} (wh_i * sh_i)}{\sum_{j=1}^{p} (wh_j * dh_j)} - \sum_{k=1}^{q} (w'h_k * sh'_k) \right)$$

$$red\_flag = 0.448275862 + \left(\frac{0.3 + 1.5 + 0 + 0 + 0 + 0 + 0 + 0}{0.3 + 1.5 + 0.7 + 0.7 + 0.3 + 0.7 + 0.6 + 0.3} - 0 \right)$$

$$red\_flag = 0.448275862 + \left(\frac{1.8}{5.1} \right)$$

$$red\_flag = 0.448275862 + 0.352941176$$

$red\_flag = 0.801217038$

We compare this result with a second threshold of value 0.6, which is retrieved after a series of tests on our limited data. If this value is greater or equal to the threshold, a red flag email will be sent to the person(s) in charge.

4.1.1 Threshold

To choose the threshold for our system, we did some testing. These tests are run using the red flag formulation and information from Figure 4.1 and Figure 4.2 for anthrax, Figure 4.3 and Figure 4.4 for botulism, Figure 4.5 and Figure 4.6 for brucellosis, Figure 4.7 and Figure 4.8 for q-fever. Some of the tests are given below:

1. Given symptoms of cattle when it is alive:
   - Shivering, trembling
   - Absence of appetite
   - Incoordination

   The result is:
   - anthrax : 0.10344827586206896
   - botulism : 0.39999999999999997
   - brucellosis : 0.0
   - q-fever : 0.0.
2. Given symptoms of cattle when it is alive:
   - Shivering, trembling
   - Paralysis
   - Incoordination

   The result is:
   - anthrax : 0.10344827586206896
   - botulism : 0.6666666666666666
   - brucellosis : 0.0
   - q-fever : 0.0.
3. Given symptoms of cattle when it is alive:
   - Shivering, trembling

- Shortness of breath
- Stupor
- Disorientation

The result is:

- anthrax : 0.4137931034482758
- botulism : 0.15789473684210525
- brucellosis : 0.0
- q-fever : 0.0.

4. Given symptoms of cattle when it is alive:
   - Shivering, trembling
   - Shortness of breath
   - Swellings on the body (e.g. ventral neck, thorax, shoulders)
   - Excitement followed by depression

   The result is:

   - anthrax : 0.48275862068965525
   - botulism : 0.16666666666666666
   - brucellosis : 0.0
   - q-fever : 0.0.

From our tests, we found that threshold = 0.45 for red flag value that uses animal symptoms gives the best result for our limited data in identifying the highly-threatening diseases. As we can see from the example above, in the first query, the given symptoms are most similar to botulism. However, these symptoms are not sufficient to indicate that botulism has occurred. Hence, we need to use a threshold that won't raise a red flag for these symptoms. Since the red flag value of these symptoms is less than 0.45, a red flag won't be raised.

For the second query, we expect that a red flag value will be raised for these symptoms. As we can see, there is symptom "Paralysis" in the given symptoms. This symptom combines with the other given symptoms are strong indicators that botulism maybe have occurred. Using threshold = 0.45, a red flag will be raised for these symptoms.

The symptoms given in the third query are not strong indicators that anthrax has occurred. These symptoms are not unique because they can be seen in other diseases. These symptoms are not sufficient to indicate the occurrence of anthrax. Hence, we expect the red flag won't be raised for these symptoms. Since the red flag value of these symptoms for anthrax is below the threshold, no red flag will be raised.

As for the fourth query, symptoms "Swellings on the body (e.g. ventral neck, thorax, shoulders)" and "Excitement followed by depression" combine with the other given symptoms can be a good indicator that anthrax has occurred. Using our threshold, a red flag will be raised for these symptoms. For human symptoms, some example will be given below.

1. Given symptoms of cattle when it is alive:
   - Shivering, trembling
   - Absence of appetite
   - Incoordination

   Given symptoms of human that has direct contact with the sick animal:
   - Difficulty swallowing

   The highest result is for botulism with the red flag value as following:
   - red flag value using animal symptoms : 0.39999999999999997
   - red flag value using animal and human symptoms : 0.5454545454545454
2. Given symptoms of cattle when it is alive:
   - Shivering, trembling
   - Absence of appetite
   - Incoordination

   Given symptoms of human that has direct contact with the sick animal:
   - Difficulty swallowing
   - Drooping eyes

   The highest result is for botulism with the red flag value as following:
   - red flag value using animal symptoms : 0.39999999999999997
   - red flag value using animal and human symptoms : 0. 0.6909090909090909

From our tests on our data, we found that threshold = 0.6 for red flag value that use animal and human symptoms in the calculation gave the best result for identifying highly-threatening diseases. In the first query, the given animal symptoms are not sufficient to determine that botulism has occurred. So we also use the given human symptoms in the calculation. In the first query, the user only gives a human symptom “Difficulty swallowing”. This symptom can be caused by many different reasons. Since this symptom is not a strong indicator and there is no other human symptom given, no red flag will be raised.

In the second query, the symptoms “Drooping eyes” and “Difficulty swallowing” are given for human symptoms. These human symptoms combine with the given animal symptoms give good indicators that botulism may have occurred. Since the red flag value is above the threshold, a red flag will be raised.

When the given animal symptoms are sufficient to indicate the occurrence of a highly-threatening disease, we don’t need to calculate the human symptoms. However, when the given animal symptoms fail to indicate the occurrence of a highly-threatening disease, the given human symptoms (if present) will be used in the calculation.

As we can see, the ability of our system to determine the occurrence of highly-threatening diseases highly depends on the thresholds’ values.

4.2 Evaluation

In this subsection, the results of running our system using several queries will be given. In the following testing, we run our system using data from tables in Figure 4.1 and Figure 4.2 for anthrax, Figure 4.3 and Figure 4.4 for botulism, Figure 4.5 and Figure 4.6 for brucellosis, and Figure 4.7 and Figure 4.8 for q-fever. A threshold = 0.45 is used to decide whether we need to raise a red flag. In the case when human symptoms are used in the calculation, a threshold = 0.6 is used.

No	Id	Symptom	Weight	Group	Stat
1	1	Found dead	0.4	1	1
2	3	Shivering, trembling	0.5	1	0
3	52	Paralysis	1	1	0
4	51	Difficulty swallowing	0.5	1	0
5	27	Absence of appetite	0.2	1	0
6	56	Incoordination	0.5	1	0
7	40	Severe breathing problems	0.3	1	0
8	2	Fever	0.8	2	0
9	13	Singeing of the hair (by a history of electrical storm)	1	2	1
10	14	Crepitating swelling	1	2	1
11	6	Bloody discharge from nose, mouth and anus	1	2	1

Figure 4.3: Table of botulism's symptoms in cattle.

No	Id	Symptom	Weight	Group	Stat
1	2	Fever	0.3	1	2
2	43	Weakness	0.3	1	2
3	44	Dizziness	0.3	1	2
4	45	Dry mouth	0.3	1	2
5	46	Nausea	0.3	1	2
6	47	Vomiting	0.3	1	2
7	40	Severe breathing problems	0.8	1	2
8	49	Double vision	0.8	1	2
9	50	Drooping eyes	0.8	1	2
10	51	Slurred speech	0.8	1	2
11	52	Paralysis	0.8	1	2

Figure 4.4: Table of botulism's symptoms in human.

Query 4.1:

- Symptoms:

- Symptoms of cattle when it is alive:
    - Shivering, trembling
    - Severe breathing problems
    - Paralysis
- Symptoms of human has direct contact with the animal:
    - Double vision
    - Severe breathing problems

  - Paralysis

- Corresponding weights:

- Symptoms of cattle when it is alive:
  - Shivering, trembling
    - anthrax: weight=0.3, type=1
    - botulism: weight=0.5, type=1
  - Severe breathing problems
    - botulism: weight=0.3, type=1
  - Paralysis
    - botulism: weight=1, type=1
- Symptoms of human has direct contact with the animal:
  - Double vision
    - botulism: weight=0.8, type=1
  - Severe breathing problems
    - anthrax: weight=0.7, type=1
    - botulism: weight=0.8, type=1
  - Paralysis
    - botulism: weight=0.8, type=1

- Calculation:

- Anthrax:

$$red\_flag = \frac{0+0.3+0+0+0+0+0+0+0+0}{0.2+0.3+0.3+0.3+0.4+0.4+0.3+0.3+0.2+0.2}$$

$$red\_flag = 0.10344827586206898$$

- Botulism

$$red\_flag = \frac{0.5+1+0+0+0+0.3}{0.5+1+0.5+0.2+0.5+0.3}$$

$$red\_flag = 0.6$$

- Brucellosis

$$red\_flag = 0.0$$

- Q-fever

$$red\_flag = 0.0$$

Here, we can see that the highest red flag value is 0.6 for botulism. Since it's higher than the 0.45, so a red flag will be raised. Since using the animal symptoms are sufficient to determine which disease may have occurred, we don't need to calculate using human symptoms.

No	Id	Symptom	Weight	Group	Stat
1	11	Abortion	0.8	1	0
2	15	Reduced fertility	0.4	1	0
3	16	Weakness in young	0.5	1	0
4	17	Retained placenta	0.5	1	0
5	18	Uterine infection	0.3	1	0
6	19	Lameness	0.3	1	0
7	20	Fetus has pneumonia	0.3	1	0
8	21	Hygromatous swellings especially of the knees (looks like a water-bag)	1.5	1	0
9	22	Abortion during < 6 months	0.1	2	0
10	23	Autolyzed fetus	0.2	2	0
11	24	Fetus is small raised, gray buff, soft lessions / wound or diffuce white areas on skin	0.2	2	0

Figure 4.5: Table of brucellosis's symptoms in cattle.

No	Id	Symptom	Weight	Group	Stat
1	42	Flu-like signs	0.2	1	2
2	48	Sweating, particularly at night	0.5	1	2
3	53	Muscular pain	0.5	1	2
4	54	Cyclic fever	0.3	1	2
5	55	Inflammation of testis	0.3	1	2

Figure 4.6: Table of brucellosis's symptoms in human.

No	Id	Symptom	Weight	Group	Stat
1	63	Sporadic abortion	0.6	1	0
2	15	Reduced fertility	0.3	1	0
3	16	Weakness in young	0.5	1	0
4	17	Retained placenta	0.5	1	0

Figure 4.7: Table of q-fever's symptoms in cattle.

No	Id	Symptom	Weight	Group	Stat
1	42	Flu-like signs	0.2	1	2
2	43	Weakness	0.3	1	2
3	57	Severe sweats	0.9	1	2
4	58	Pneumonia	0.8	1	2
5	61	Retrobulbar headache	0.6	1	2
6	62	Inflammation of the placenta	0.3	1	2
7	59	Productive Cough	0.8	2	2
8	60	Chest pain	0.8	1	2

Figure 4.8: Table of q-fever's symptoms in human.

Query 4.2:

- Symptoms:
- Symptoms of cattle when it is alive:
    - Fever
    - Shivering, trembling
    - Shortness of breath
    - Swellings on the body (e.g. ventral neck, thorax, shoulders)
    - Excitement followed by depression
- Symptoms of human has direct contact with the animal:
    - Unknown
- Corresponding weights:
- Symptoms of cattle when it is alive:
    - Fever
        - anthrax: weight=0.2, type=1
        - botulism: weight=0.8, type=2
    - Shivering, trembling
        - anthrax: weight=0.3, type=1
        - botulism: weight=0.5, type=1
    - Shortness of breath
        - anthrax: weight=0.3, type=1
    - Swellings on the body (e.g. ventral neck, thorax, shoulders)
        - anthrax: weight=0.4, type=1

  - Excitement followed by depression
    - anthrax: weight=0.4, type=1

- Calculation:
- Anthrax:

$$red\_flag = \frac{0.2+0.3+0.3+0+0.4+0.4+0+0+0+0}{0.2+0.3+0.3+0.3+0.4+0.4+0.3+0.3+0.2+0.2}$$

$$red\_flag = 0.5517241379310346$$

- Botulism

$$red\_flag = \frac{0.5+0+0+0+0+0}{0.5+1+0.5+0.2+0.5+0.3} - 0.8$$

$$red\_flag = -0.6333333333333334$$

- Brucellosis

$$red\_flag = 0.0$$

- Q-fever

$$red\_flag = 0.0$$

Here, we can see that the highest red flag value is 0.5517241379310346 for anthrax. Since it's higher than the 0.45, so a red flag will be raised.

Query 4.3:

- Symptoms:
- Symptoms of cattle when it is alive:
  - Abortion
  - Retained placenta
  - Lameness
- Symptoms of human has direct contact with the animal:
  - Sweating, particularly at night
  - Muscular pain
  - Cyclic fever
- Corresponding weights:
- Symptoms of cattle when it is alive:
  - Abortion

- anthrax: weight=0.2, type=1
- brucellosis: weight=0.8, type=1

o Retained placenta

- brucellosis: weight=0.5, type=1
- q-fever: weight=0.5, type=1

o Lameness

- brucellosis: weight=0.3, type=1

- Calculation:

- Anthrax:

$$red\_flag = \frac{0.2+0+0+0+0+0+0+0+0+0}{0.2+0.3+0.3+0.3+0.4+0.4+0.3+0.3+0.2+0.2}$$

$$red\_flag = 0.06896551724137931$$

- Botulism

$$red\_flag = \frac{0}{0.5+1+0.5+0.2+0.5+0.3}$$

$$red\_flag = 0.0$$

- Brucellosis

$$red\_flag = \frac{0.8+0+0+0.5+0+0.3+0+0}{0.8+0.4+0.5+0.5+0.3+0.3+0.3+1.5}$$

$$red\_flag = 0.34782608695652173$$

- Q-fever

$$red\_flag = \frac{0+0+0+0.5}{0.6+0.3+0.5+0.5}$$

$$red\_flag = 0.2631578947368421$$

Here, we can see that the highest red flag value is 0.34782608695652173 for brucellosis. Since it's below 0.45, so we need to use the given human symptoms to recalculate the red flag value.

- Corresponding weights:

- Symptoms of human has direct contact with the animal:

o Sweating, particularly at night

- brucellosis: weight=0.5 type=1

o Muscular pain

        - brucellosis: weight=0.5 type=1
    - Cyclic fever
        - brucellosis: weight=0.3, type=1

- Calculation:
- Anthrax:

$$red\_flag = \frac{0.2+0+0+0+0+0+0+0+0+0}{0.2+0.3+0.3+0.3+0.4+0.4+0.3+0.3+0.2+0.2} + \frac{0}{0.3+1.5+0.7+0.7+0.3+0.7+0.6+0.3}$$

$$red\_flag = 0.06896551724137931$$

- Botulism

$$red\_flag = \frac{0}{0.5+1+0.5+0.2+0.5+0.3} + \frac{0}{0.3+0.3+0.3+0.3+0.3+0.3+0.8+0.8+0.8+0.8+0.8}$$

$$red\_flag = 0.0$$

- Brucellosis

$$red\_flag = \frac{0.8+0+0+0.5+0+0.3+0+0}{0.8+0.4+0.5+0.5+0.3+0.3+0.3+1.5} + \frac{0+0.5+0.5+0.3+0}{0.2+0.5+0.5+0.3+0.3}$$

$$red\_flag = 1.0700483091787438$$

- Q-fever

$$red\_flag = \frac{0+0+0+0.5}{0.6+0.3+0.5+0.5} + \frac{0}{0.2+0.3+0.9+0.8+0.6+0.3+0.8+0.8}$$

$$red\_flag = 0.2631578947368421$$

In this query, the highest red flag value is 1.0700483091787438 for brucellosis. Since it's above 0.6, the second threshold, a red flag is raised and an email contains information about the user, the symptoms, the suspected disease is sent to the person(s) in charge. Here, we can see that the information about human symptoms can be useful when the given animal symptoms are not sufficient to predict whether a highly-threatening disease has occurred.

Query 4.4:

- Symptoms:
- Symptoms of cattle when it is alive:
    - Shivering, trembling

    - Absence of appetite
    - Incoordination
    - Lameness

- Corresponding weights:
- Symptoms of cattle when it is alive:
    - Shivering, trembling
        - anthrax: weight=0.3, type=1
        - botulism: weight=0.5, type=1
    - Absence of appetite
        - botulism: weight=0.2, type=1
    - Incoordination
        - botulism: weight=0.5, type=1
    - Lameness
        - brucellosis: weight=0.3, type=1
- Calculation:
- Anthrax:

$$red\_flag = \frac{0+0.3+0+0+0+0+0+0+0+0}{0.2+0.3+0.3+0.3+0.4+0.4+0.3+0.3+0.2+0.2}$$

$$red\_flag = 0.10344827586206896$$

- Botulism

$$red\_flag = \frac{0.5+0+0+0.2+0.5+0}{0.5+1+0.5+0.2+0.5+0.3}$$

$$red\_flag = 0.39999999999999997$$

- Brucellosis

$$red\_flag = \frac{0+0+0+0+0+0.3+0+0}{0.8+0.4+0.5+0.5+0.3+0.3+0.3+1.5}$$

$$red\_flag = 0.06521739130434782$$

- Q-fever

$$red\_flag = \frac{0}{0.6+0.3+0.5+0.5}$$

$$red\_flag = 0.0$$

Here, we can see that the highest red flag value is 0.39999999999999997 for botulism. There is no red flag will be raised because the red flag value is below the first threshold 0.45 and there is no human symptom is given. The existing symptoms are not sufficient to decide that botulism may have occurred and there is no human symptoms given to support the suspicion that botulism has occurred.

Query 4.5:

- Symptoms:
- Symptoms of cattle when it is alive:
    - Shivering, trembling
    - Excitement followed by depression
    - Abortion
- Symptoms of cattle when it is dead:
    - Found dead
    - Bloody discharge from nose, mouth and anus
    - Crepitating swelling
- Corresponding weights:
- Symptoms of cattle when it is alive:
    - Shivering, trembling
        - anthrax: weight=0.3, type=1
        - botulism: weight=0.5, type=1
    - Excitement followed by depression
        - anthrax: weight=0.4, type=1
    - Abortion
        - anthrax: weight=0.2, type=1
        - brucellosis: weight=0.8, type=1
- Symptoms of cattle when it is dead:
    - Found dead
        - anthrax: weight=0.8, type=1
        - botulism: weight=0.4, type=1

- Bloody discharge from nose, mouth and anus
  - anthrax: weight=1, type=1
  - botulism: weight=1, type=2
- Crepitating swelling
  - anthrax: weight=1, type=2
  - botulism: weight=1, type=2

- Calculation:

- Anthrax:
  - When the cattle is alive

$$red\_flag = \frac{0+0.3+0+0+0+0.4+0+0+0.2+0}{0.2+0.3+0.3+0.3+0.4+0.4+0.3+0.3+0.2+0.2}$$

$$red\_flag = 0.31034482758620685$$

  - When the cattle is dead

$$red\_flag = \frac{0.8+1}{0.8+1} - 1$$

$$red\_flag = 0.0$$

  - Since the red flag value of symptoms when the cattle are alive is higher then when the cattle is dead, so we use the first red flag value, that is 0.31034482758620685, as the red flag value of anthrax.

- Botulism
  - When the cattle is alive

$$red\_flag = \frac{0.5+0+0+0+0+0}{0.5+1+0.5+0.2+0.5+0.3}$$

$$red\_flag = 0.16666666666666666$$

  - When the cattle is dead

$$red\_flag = \frac{0.4}{0.4} - (1+1)$$

$$red\_flag = -1.0$$

  - Since the red flag value of symptoms when the cattle are alive is higher then when the cattle is dead, so we use the first red flag value, that is 0.16666666666666666, as the red flag value of botulism.

- Brucellosis

$$red\_flag = \frac{0.8+0+0+0+0+0.3+0+0}{0.8+0.4+0.5+0.5+0.3+0.3+0.3+1.5}$$

$$red\_flag = 0.17391304347826086$$

- Q-fever

$$red\_flag = \frac{0}{0.6+0.3+0.5+0.5}$$

$$red\_flag = 0.0$$

Here, we can see that the highest red flag value is 0.31034482758620685 for anthrax. Since this value is below the first threshold and no human symptoms are given, no red flag is raised.

For the symptoms when the animal is dead, they seem similar to anthrax's symptoms. However, symptom "Crepitating swelling shows that it is blackleg, not anthrax. The fact that the user chooses this symptom, take away our suspicion that it is anthrax. By choosing this symptom, the value is reduced to 0. Since we take the highest red flag value, the red flag value of symptoms when the animal is alive is selected.

Query 4.6:

- Symptoms:
- Symptoms of cattle when it is alive:
    - Retained placenta
    - Sporadic abortion
- Symptoms of human has direct contact with the animal:
    - Flu-like signs
    - Severe sweats
    - Pneumonia
- Corresponding weights:
- Symptoms of cattle when it is alive:
    - Retained placenta
        - brucellosis: weight=0.5, type=1
        - q-fever: weight=0.5, type=1

  - Sporadic Abortion
    - q-fever: weight=0.6, type=1

- Calculation:

- Anthrax:

$$red\_flag = \frac{0}{0.2+0.3+0.3+0.3+0.4+0.4+0.3+0.3+0.2+0.2}$$

$$red\_flag = 0.0$$

- Botulism

$$red\_flag = \frac{0}{0.5+1+0.5+0.2+0.5+0.3}$$

$$red\_flag = 0.0$$

- Brucellosis

$$red\_flag = \frac{0+0+0+0.5+0+0+0+0}{0.8+0.4+0.5+0.5+0.3+0.3+0.3+1.5}$$

$$red\_flag = 0.10869565217391303$$

- Q-fever

$$red\_flag = \frac{0.6+0+0+0.5}{0.6+0.3+0.5+0.5}$$

$$red\_flag = 0.5789473684210527$$

Here, we can see that the highest red flag value is 0.5789473684210527 for q-fever. Since this value is above 0.45, the first threshold, a red flag value is raised to inform the person(s) in charge that q-fever may have occurred. Here, no further calculation using human symptoms need to be performed.

Query 4.7:

- Symptoms:

- Symptoms of cattle when it is alive:
  - Disorientation
  - Reduced Fertility
  - Lameness

- Corresponding weights:

- Symptoms of cattle when it is alive:

- Disorientation
  - anthrax: weight=0.3, type=1
- Reduced Fertility
  - brucellosis: weight=0.4, type=1
  - q-fever: weight=0.3, type=1
- Lameness
  - brucellosis: weight=0.3, type=1

- Calculation:

- Anthrax:

$$red\_flag = \frac{0+0+0+0+0+0+0+0.3+0+0}{0.2+0.3+0.3+0.3+0.4+0.4+0.3+0.3+0.2+0.2}$$

$$red\_flag = 0.10344827586206896$$

- Botulism

$$red\_flag = \frac{0}{0.5+1+0.5+0.2+0.5+0.3}$$

$$red\_flag = 0.0$$

- Brucellosis

$$red\_flag = \frac{0+0.4+0+0+0+0.3+0+0}{0.8+0.4+0.5+0.5+0.3+0.3+0.3+1.5}$$

$$red\_flag = 0.15217391304347827$$

- Q-fever

$$red\_flag = \frac{0+0.3+0+0}{0.6+0.3+0.5+0.5}$$

$$red\_flag = 0.15789473684210525$$

Here, we can see that the highest red flag value is 0.15789473684210525 for q-fever. Since this value is less than 0.45, the first threshold, a red flag value is not raised.

Query 4.8:

- Symptoms:

- Symptoms of cattle when it is alive:
  - Reduced fertility
  - Retained placenta

    - Abortion during < 6 months

- Corresponding weights:

- Symptoms of cattle when it is alive:
    - Reduced Fertility
        - brucellosis: weight=0.4, type=1
        - q-fever: weight=0.3, type=1
    - Retained placenta
        - brucellosis: weight=0.5, type=1
        - q-fever: weight=0.5, type=1
    - Abortion during < 6 months
        - brucellosis: weight=0.1, type=2

- Calculation:

- Anthrax:

$$red\_flag = \frac{0}{0.2+0.3+0.3+0.3+0.4+0.4+0.3+0.3+0.2+0.2}$$

$$red\_flag = 0.0$$

- Botulism

$$red\_flag = \frac{0}{0.5+1+0.5+0.2+0.5+0.3}$$

$$red\_flag = 0.0$$

- Brucellosis

$$red\_flag = \frac{0+0.4+0+0.5+0+0+0+0}{0.8+0.4+0.5+0.5+0.3+0.3+0.3+1.5} - 0.1$$

$$red\_flag = 0.09565217391304348$$

- Q-fever

$$red\_flag = \frac{0+0.3+0+0.5}{0.6+0.3+0.5+0.5}$$

$$red\_flag = 0.42105263157894735$$

The highest red flag value for this query is 0.42105263157894735 for q-fever. Since this value is less than 0.45, the first threshold, a red flag value is not raised.

We can see from those queries and their results that the value of weights we used is one of the keys that determine how effective our system in identifying the occurrence of a highly-threatening disease. The main challenge is to decide which symptoms give more valuable information compare to the others. Being able to identify unique symptoms for each disease accurately will improve the accuracy of our system in identifying a disease that occurs. In addition, most of the times the given symptoms are only partial and from these symptoms, sometimes the user missed the symptoms that give the most valuable information. Hence, in assigning the weight of each symptom of a disease, we also need to consider which symptoms are usually noticed by a user and which symptoms are usually left behind. Note that the weights have been set using information from the literature on the various diseases and animal species.

From the tests we run, we can see that in Query 4.1, Query 4.2, Query 4.3, and Query 4.6 where the red flag are raised, the symptoms given by the user contain information that strongly indicates the occurrence of highly-threatening diseases. As example in Query 4.1, symptom “Paralysis” which is one of the unique characteristic of botulism is given by the user with the other botulism’s symptoms. In Query 4.2, the user gives symptoms “Excitement followed by depression” and “Swellings on the body (e.g. ventral neck, thorax, shoulders)” together with other anthrax’s symptoms. These occurrences of these symptoms strengthened the prediction that anthrax has occurred. Together with other given anthrax’s symptoms, they become a strong indicator of the occurrence of anthrax. In Query 4.6, the user specifies symptom “Sporadic abortion” which is one of the characteristic of q-fever that differentiates this disease with the others. Together with the symptom “retained placenta”, they are a good indicator that q-fever may have occurred.

In Query 4.3 where the animal’s symptoms given by the user are not quite strong indicators, the user also gives the symptoms of human in direct contact with the animal. When the animal symptoms are combined with the human symptoms, they become very good indicators that brucellosis may have occurred. In Queries 4.4, 4.5, 4.7, and 4.8, there is no red flag raised because the given symptoms are not sufficient in indicating that a highly-threatening disease has occurred.

As we can see from the result of our queries, using our setting for the weights and the thresholds as described before, our system is able to identify the occurrences of highly-threatening diseases for the data that we have compiled to this point. It raises a red flag when the given symptoms resemble the occurrence of a highly-threatening disease. By setting our weights and threshold with the right value, the occurrences of false positive and false negative red flag will be reduced. We realize that false negative and false positive still may occur, because sometime users do not give sufficient descriptions of their animal symptoms.

The implementation of our system will be discussed in chapter 5.

5. IMPLEMENTATION

In Chapters 3 and 4, we have talked about our approach in providing a user with some information about caring for animals with certain symptoms while analyzing whether there is a chance that any highly-threatening disease has occurred. This chapter describes the implementation of our system.

5.1 User Interface

As mentioned in Chapter 3, a user interacts with our system using a web-based interface. This web-based interface has been implemented using Java Server Pages (JSP) within the Apache Struts framework. Apache Struts is an open-source web application framework for developing Java EE web applications which encourages developers to adopt a model-view-controller (MVC) architecture such that it is easier to maintain [2]. We use Apache Tomcat 5.0 as our application server.

5.1.1 User Authentication

Through our web pages, a user can login into our system from the login page and start searching for information about symptoms that his/her animal show. Figure 5.1 shows an existing user trying to login into our system. He/she enters his/her username, password, group and clicks the submit button. Once the system finds a match between the username, password and group given by the user and the user information stored in database, the user is logged in into our system, otherwise an error message that the information given is invalid will be shown. Once the user enters our system, the user can edit his/her account information or start searching information about his/her animal symptoms.

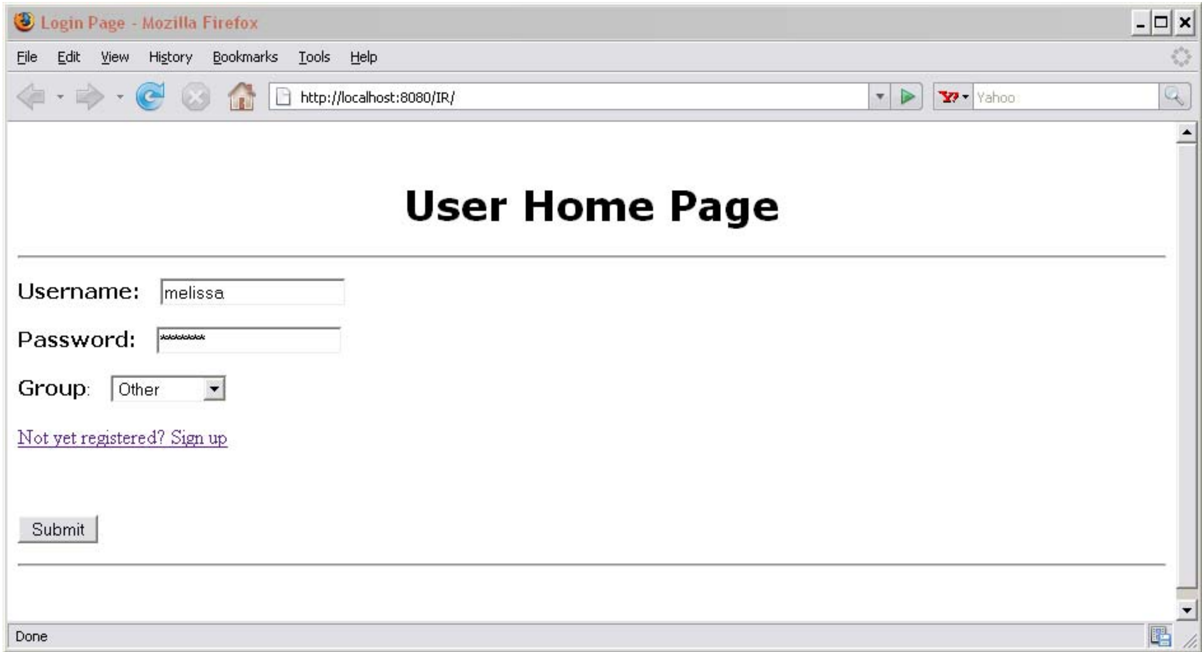

Figure 5.1: User login page.

5.1.2 User Registration

A new user who is interested in our system can register through our web-based interface by clicking a link "Not yet registered? Sign up" on the login page shown in Figure 5.1. This link will lead him/her to the registration page where the user will need to provide information such as a username that he/she wants to use, a password, the group where the user belongs to, his/her name, address, city, zip code, phone, email address, and the university where he/she comes from as shown in Figure 5.2. The user should choose a username and password that have more than 6 characters each. After the user submits the information, a confirmation page (Figure 5.3) will be shown to inform that the information has been successfully saved. From here, the user can choose whether he/she wants to continue going to the main page by clicking "Continue to the main page" button or going back to the login page by clicking "Back to login page" button. By going to the main page, the user will be able to start searching information about his/her animal's symptoms or edit the user account if he/she wishes to do so.

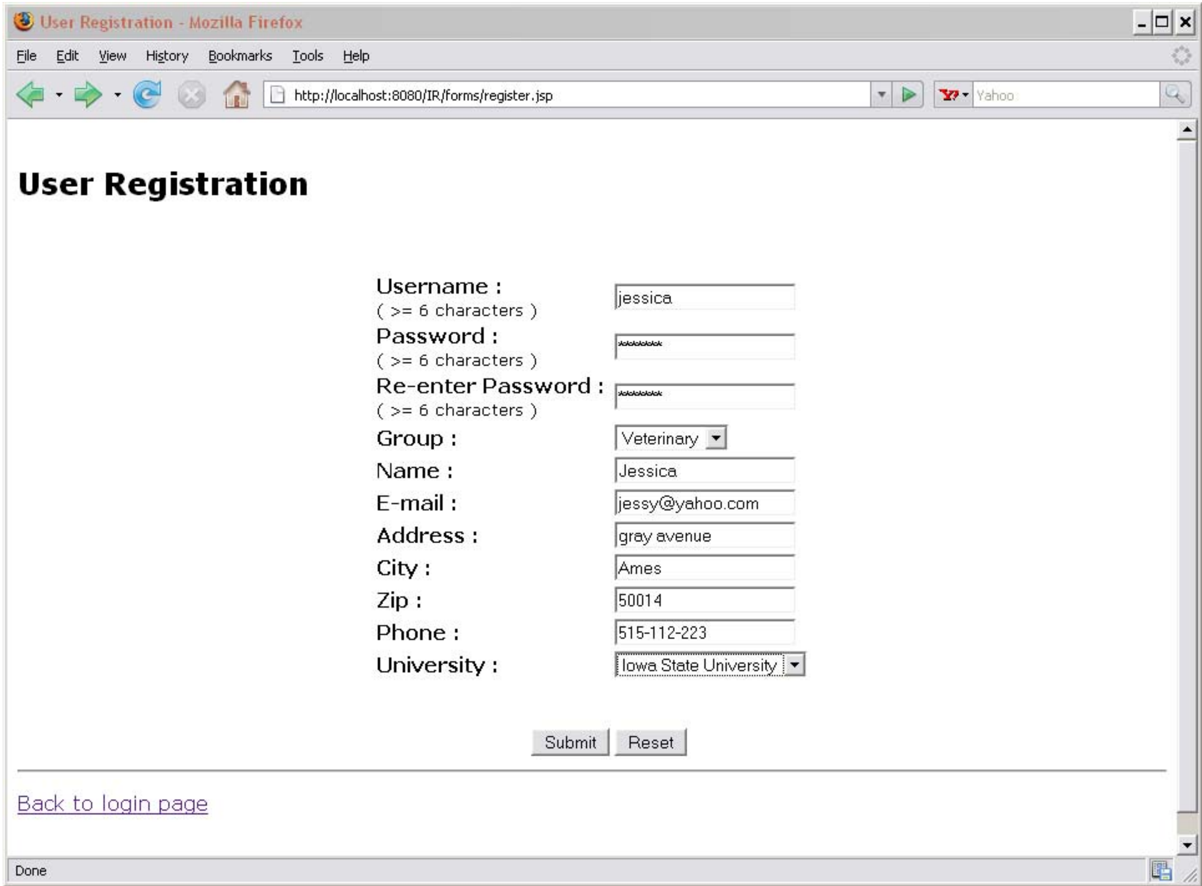

Figure 5.2: User Registration page.

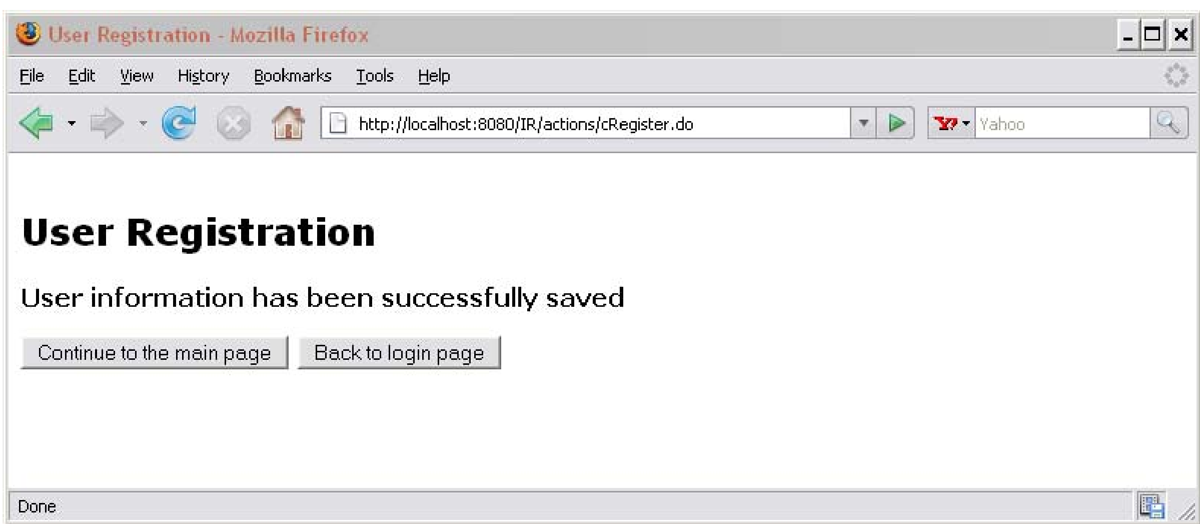

Figure 5.3: User Registration Confirmation page.

5.1.3 User Account Information

Once a user enters our system, the user can view and change his/her account information by clicking the menu "Search Animal Information". It will lead the user to a web page to edit the user information (Figure 5.4) where the user can also see all of his/her current account information. The user can also change his/her password as well from this web page by checking the checkbox to change password. By doing this, the user will be asked to input the new password as well. Once all the information needed has been changed, the user need to clicks submit to overwrite the old information. A message will be shown to inform the user whether there is any information missing or the change has been completed successfully.

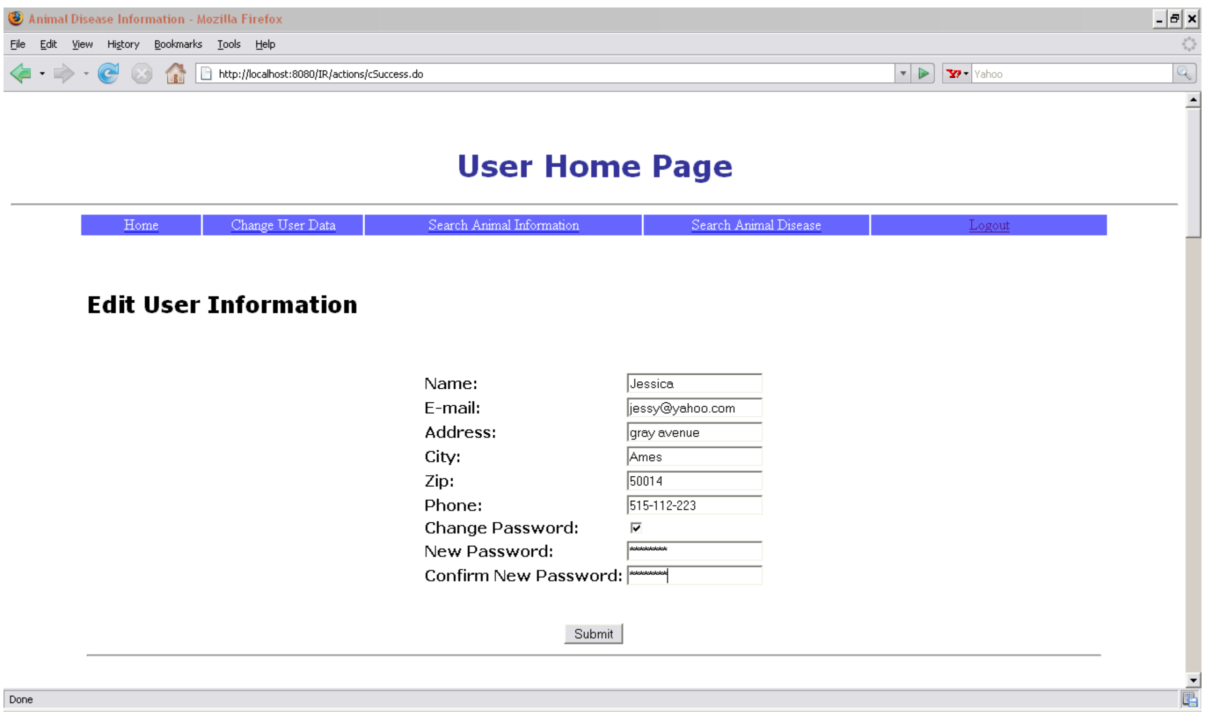

Figure 5.4: Web page to edit user information.

5.1.4 Searching Animal Disease Information

To search for some information about an animal disease, a user can click "Search Animal Disease" from the menu. A search animal information web page (Figure 5.5) will appear with the list of animal species, list of animal symptoms when it is alive, when it is dead, and list of symptoms of human that has direct contact with the animal.

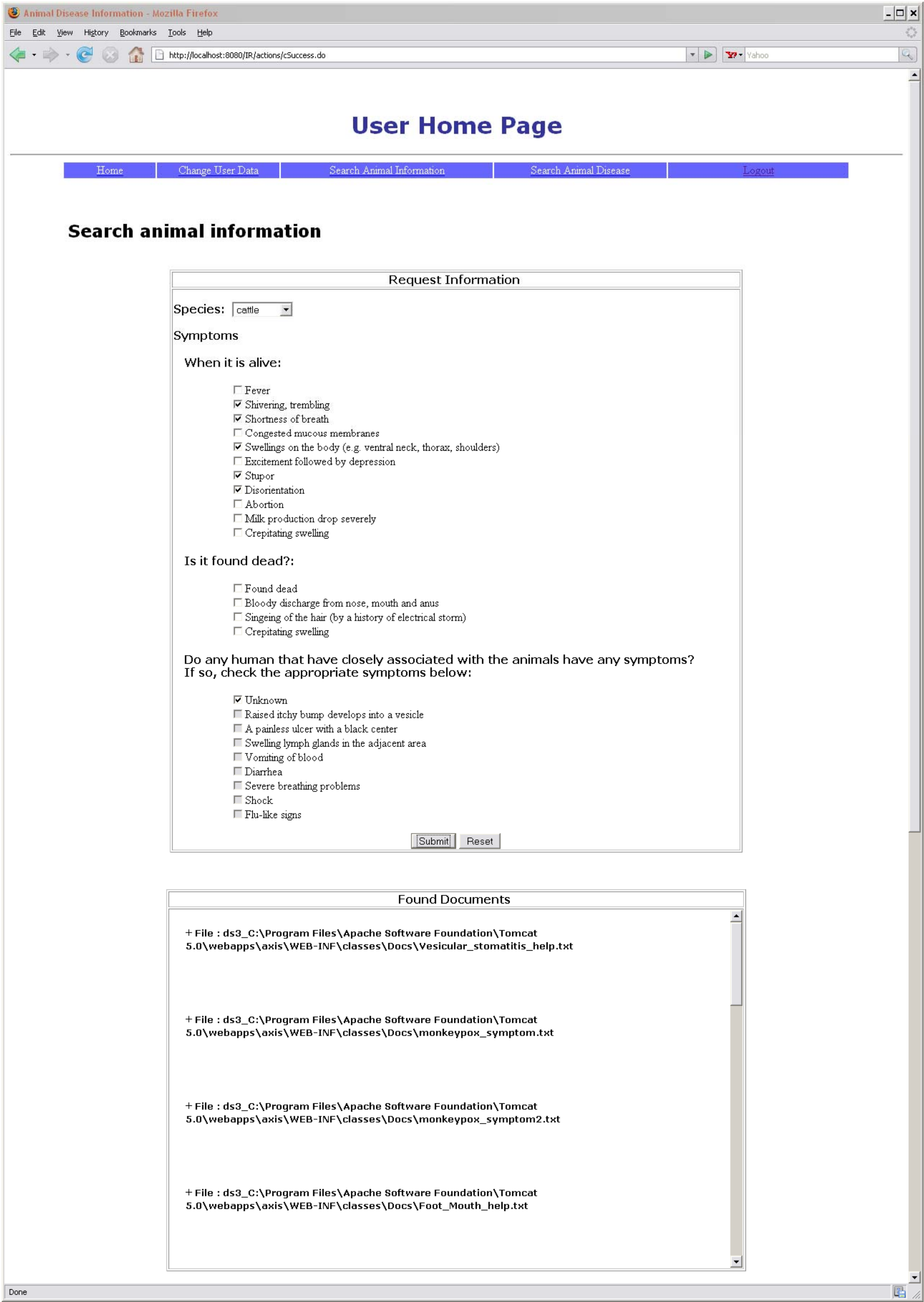

Figure 5.5: Web page showing the search for animal disease information and the results returned for the user query.

Inside the table “Request Information”, the user needs to specify the species of his/her animal, the symptoms his/her animal shows and the symptoms of any human in contact with the animal if there is any and click submit. First, the system will check which data sources that the user has access to and contain the information that the user looking for. The system will check the university where the user comes from and select data sources owned by that university. From this data sources, the system checks which data sources have diseases information of the animal species specified by the user. After that, using the species name and list of symptoms given by the user as keywords, the system searches for documents in the data sources that have been selected previously.

In the meantime, the system initiates the *red flag analyzer* process to analyze the species and symptoms given by the user to decide whether there is any red flag need to be raised. Then, the red flag analyzer calculates the red flag value of highly-threatening diseases using these diseases information stored in the *internal object oriented database* and the information given by the user as described in Chapter 4. If the red flag analyzer finds that there are high similarities between symptoms given by the user and symptoms of a highly-threatening disease, then a red flag is raised. An email contains information about the user, the disease, the symptoms, and an attachment of the most relevant document will be sent to the person(s) in charge.

Once the system finishes retrieving all relevant documents from the data sources, the result are shown to the user inside the table “Found Documents”. To read the content of these documents, the user can click the document name such that the description of the document will be shown followed by a link “Read the Content” that leads to another page contains the document’s content.

5.2 Data Sources

As describes before, our system use several data sources contain documents about animal disease. These data sources reside in different sites, so we need some way to retrieve these documents. In our system, we develop a web service for each data source using Apache Axis 1.4 which implements SOAP (Simple Object Access Protocol) protocol. The main advantage of using SOAP protocol is firewall friendly because it is an XML-based wire protocol [16]. We use Axis to write and deploy web services on the servers and to write the clients. Using axis to write the clients, we only need to make method calls on the web service object as if it were a local object [9].

We will talk about the implementation of our system's data source in Section 5.2.1. We can see the general structure of the data sources in Figure 3.4.

5.2.1 Implementation

As described in Chapter 3, a *mediator* sends a user query to the *wrapper* of a data source. This *wrapper* is implemented by class **DataSource_1** for the first data source, **DataSource_2** for the second data source, and **DataSource_3** for the third data source. To process the query, the *wrapper* creates objects of data source's *LIV*. If it needs to index the data source's documents, it initiates the indexing process by calling method *doIndexing()*. This indexing process is performed daily. If it needs to search documents relevant to the user query, this process will be initiated by calling method *doMatching()*. This data source's *LIV* is implemented by class **SoapClient_1** for the first data source, **SoapClient_2** for the second data source, and **SoapClient_3** for the third data source. Each method in the data source's *LIV* corresponds to a data source's tool that resides in different site. These tools are implemented by methods of class **LIV**. The data source's *LIV* allows the wrapper to use the data source's tools that reside in different sites, as if it is local. **LIV** is the class that implements the SOAP server and the data source's *LIV* implements the SOAP client.

The data source's LIV interacts with methods in **LIV**. The method in **LIV** will create an object of **Doc_LIV** if the method is related to the indexing process. If the method is related to the matching process of documents and the user query, it will create an object of **Matching_LIV**. Then, this method will initiate the corresponding process. Once the process is completed and the result is obtained, this result will be sent back to the corresponding data source's *LIV*. Subsequently, the result will be returned to the data source's *wrapper*. Then, the *wrapper* will return it to the *mediator*.

5.3 Internal Object Oriented Database

To store our data such as users information, data sources, diseases, etc, in *internal object oriented database*, we use a relational SQL Database Management System MySQL, and Hibernate, an object-relational mapping library for Java which provides framework to map an object-oriented domain model to a traditional relational database [11]. MySQL is used to implement the relational database in the *internal object oriented database* back end.
Hibernate handles the mapping from Java classes to database tables and provides data query and retrieval facilities. A high-level view of the Hibernate architecture is shown in Figure 5.6.

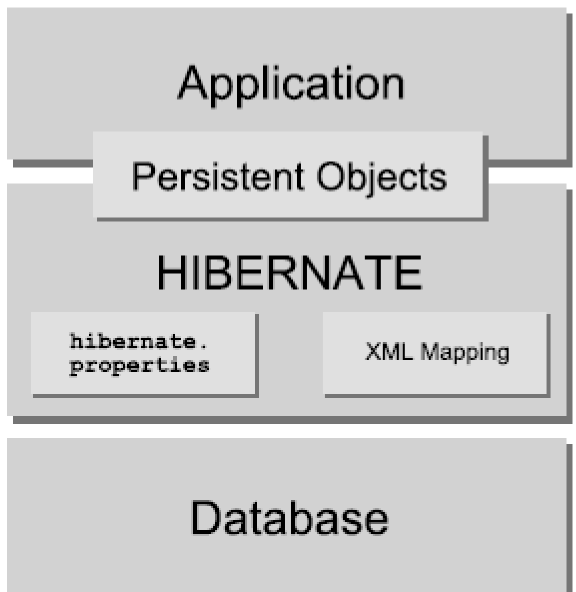

Figure 5.6: Hibernate architecture.

From this figure, we can see that Hibernate provides persistent services (and persistent objects) to the application by using the database and configuration data. As the intermediary of our application and the database, Hibernate allows the application (written as Java code) to retrieve data from and store data to the database as a Java object. This Java object is represented as a JavaBean class, a Plain Old Java Object (POJO) that has no-argument constructor, serializable, has getter and setter methods to allow access to the object's properties.

To use Hibernate, first we should create a JavaBean class for each table in the database. The methods in the class should be mapped to the table's columns. Then, we need to create

XML mapping file to tell Hibernate about which table in the database to access and columns in that table to use.

To allow our application to use Hibernate to load and store Java objects into the *internal object oriented database* back end, we create JavaBean classes *User*, *University*, *Symptoms*, *Species*, *Diseases*, and *DataSource* along with their corresponding XML mapping files. Each of these classes corresponds to a table in the *internal object oriented database* back end (Figure 3.3). Table *DataSource_Description* is created by Hibernate from the many-to-many relationship between table *DataSource* and table *Species*. Hibernate creates table *Disease_Species_Symptom* from the ternary relationship between table *Diseases*, *Symptoms*, and *Species*.

In Figure 5.7, a Java class **Symptoms** and its XML mapping file is created for table *Symptoms*. As we can see from this example, class **Symptoms** (Figure 5.7a) has attributes *id* (Integer) and *name* (String). In the mapping file (Figure 5.7b), inside the “class” element, the Java class **Symptoms** is mapped to table *Symptoms*. Following this, each attribute of class **Symptoms** is mapped to a column in table *Symptoms*. The “id” element declares the identifier property. The value of attribute “column” in the “id” element, in this case *symptom_id*, will be the primary key of table *Symptoms*.

```
package database.src;

public class Symptoms {

    private Integer id;

    private String name;

    public Integer getId() {
        return id;
    }

    public void setId(Integer id) {
        this.id = id;
    }

    public String getName() {
        return name;
    }

    public void setName(String name) {
        this.name = name;
    }

}
```

Figure 5.7a: Symptoms as a Java class.

```
<?xml version="1.0"?>
<!DOCTYPE hibernate-mapping PUBLIC "-//Hibernate/Hibernate Mapping DTD
3.0//EN" "http://hibernate.sourceforge.net/hibernate-mapping-3.0.dtd">

<hibernate-mapping>
  <class name="database.src.Symptoms" table="Symptoms">
    <id name="id" column="symptom_id">
      <generator class="native"/>
    </id>
    <property name="name"/>
  </class>
</hibernate-mapping>
```

Figure 5.7b: Hibernate mapping file for class Symptoms.

6. CONCLUSION AND FUTURE WORKS

Bioterrorism has become a major and very feared threat, particularly with the mass production of biological weapons by several countries and the fear of its usage in future conflicts, especially by radical groups. Therefore, a lot of research has been conducted in bioterrorism surveillance to identify any potential threat at the earliest time. Most bioterrorism surveillance systems try to identify threats by continuously monitoring public health. Based on the fact that most bio-weapons are likely to be animal diseases, there is another category of systems that focus on certain animal species with the expectation that it will allow earlier identification of the threats. The truth is the susceptibility of animals toward different diseases really varies. By focusing on certain species of animal, there will always be some other diseases that are missed and won't be detected if they occur. Using our system, we have taken a preliminary step towards detecting highly-threatening diseases in a variety of animal species using the disease information stored in the *internal object oriented database*. Whenever a user of our system is searching for literatures by specifying the symptoms of the animal and the symptoms of a human who experienced direct contact with the animal, our system will automatically analyze and attempt to detect the existence of highly-threatening diseases. In a disease is detected, a red flag is raised and people in charged are properly informed. This way, instead of waiting for a certain period of time to check whether the occurrences of a certain disease are above the normal frequency, in order to conclude the occurrence of highly-threatening disease, our system would allow the identification of such a disease at a much earlier time.

Currently, all the weights for symptoms of each disease in an animal species are assigned manually from the literature. In the future, it would be helpful if this could be done automatically by examining larger sets of literature. For current red flag formulation, there may be another additional type of symptom that can be considered to be added as well to improve the accuracy of identifying a highly-threatening disease occurrence. In addition, more data sources especially from pet clinics could be integrated and a tool to analyze this data could be added as well. Our matching functions need a great deal more analysis and this analysis will be a critical part of any future work on this project

REFERENCES

[1] ”Animal Diseases Considered Bioterrorist (BT) Threats,” *Veterinary Public Health and Rabies Control Program, County of Los Angeles*, 26 Aug. 2002; http://lapublichealth.org/wwwfiles/ph/dcp/vet/btdiseas.pdf

[2] “Apache Struts,” *Wikipedia*, 2007; http://en.wikipedia.org/wiki/Apache_Struts

[3] A.S. Khan et al., “Biological and Chemical Terrorism:Strategic Plan for Preparedness and Response,” *MMWR*, 21 Apr. 2000; http://www.cdc.gov/mmwr/preview/mmwrhtml/rr4904a1.htm

[4] C. Holmes, “Why Bioterrorism is America’s Greatest Threat,” 12 July 2003; http://baltimorechronicle.com/jul03_bioterrorism.shtml.

[5] C. Langston, “Postexposure Management and Treatment of Anthrax in Dogs — Executive Councils of the American Academy of Veterinary Pharmacology and Therapeutics and the American College of Veterinary Clinical Pharmacology,” AAPS Journal, vol. 07, no. 02; 22 Sept. 2005

[6] D.C. Blood, O.M. Radostits, and J.A. Henderson, *Veterinary Medicine: A Textbook of the Diseases of Cattle, Sheep, Pigs, Goats and Horses,* Bailliere Tindall, 1983

[7] “Disease Factsheets,” *the Center for Food Security & Public Health,* 2008; http://www.cfsph.iastate.edu/DiseaseInfo/factsheets.htm

[8] D.J. Berndt, A.R. Hevner, and J. Studnicki, *Bioterrorism Surveillance with Real-Time Data Warehousing*, LNCS 2665, Springer Berlin / Heidelberg, 2003.

[9] “FrontPage/Axis/AxisGeneral,” *Ws Wiki*, 16 Dec. 2005; http://wiki.apache.org/ws/FrontPage/Axis/AxisGeneral

[10] G. Moore et al., “A Nationwide Companion Animal Syndromic Surveillance System for Detection of Infectious and Environmental Hazards,” *Epidemiology, Purdue University Department of Veterinary Pathobiology*, 2003; http://www.vet.purdue.edu/epi/Banfield_Purdue_Surveillance.doc

[11] “Hibernate (Java),” *Wikipedia*, 2008; http://en.wikipedia.org/wiki/Hibernate_(Java)

[12] “HIBERNATE - Relational Persistence for Idiomatic Java”, 2007, http://www.hibernate.org/hib_docs/v3/reference/en/html/

[13] P. Rabinowitz et al., "Animals as Sentinels of Bioterrorism Agents," *Emerging Infectious Diseases*, Apr. 2006; http://www.cdc.gov/ncidod/EID/vol12no04/05-1120.htm

[14] "SQL Tutorial: Relational Database Basics," *Webucator*, 2008; http://www.learn-sql-tutorial.com/Introduction.php

[15] "Syndromic Surveillance: An Effective Tool for Detecting Bioterrorism?," *RAND Corporation*, 2004; http://www.rand.org/pubs/research_briefs/RB9042/index1.html

[16] T. Modi, "Axis: The next generation of Apache SOAP," *JavaWorld*, 25 Jan. 2002; http://www.javaworld.com/javaworld/jw-01-2002/jw-0125-axis.html

APPENDIX A

Appendix A contains all the information about the highly-threatening diseases. This information is used to identify whether there is any of these diseases has occurred from the user query.

No	Id	Symptom	Weight	Group	Stat
1	1	Found dead	0.8	1	1
2	2	Fever	0.2	1	0
3	4	Shortness of breath	0.5	1	0
4	7	Swellings on the body (e.g. ventral neck, thorax, shoulders)	1.5	1	0
5	41	Shock	0.3	1	0
6	6	Bloody discharge from nose, mouth and anus	1	1	1
7	13	Singeing of the hair (by a history of electrical storm)	0.8	2	1

Figure 1: Table of anthrax's symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	43	Weakness	0.8	1	0
2	3	Shivering, trembling	0.8	1	0
3	51	Difficulty swallowing	0.8	1	0
4	4	Shortness of breath	0.3	1	0
5	52	Paralysis	1.5	1	0

Figure 2: Table of botulism's symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	2	Fever	0.4	1	0
2	91	Enlargement of lymph nodes	0.8	1	0

Figure 3: Table of plague's symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	42	Flu-like signs	0.3	1	2
2	91	Enlargement of lymph nodes	1.5	1	2
3	58	Pneumonia	0.8	1	2
4	40	Severe breathing problems	1.5	1	2
5	92	Cough with bloody sputum	0.8	1	2

Figure 4: Table of plague's symptoms in human.

No	Id	Symptom	Weight	Group	Stat
1	2	Fever	0.2	1	0
2	9	Stupor	0.5	1	0
3	27	Absence of appetite	0.4	1	0
4	94	Stiffness	0.4	1	0
5	93	Abnormally rapid heart rate	0.8	1	0
6	95	Very rapid respiration	1	1	0

Figure 5: Table of tularemia's symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	42	Flu-like signs	0.3	1	2
2	96	Exhaustion	0.3	1	2
3	97	Ulceration	0.5	1	2
4	91	Enlargement of lymph nodes	0.4	1	2
5	60	Chest pain	0.5	1	2
6	98	Abdominal pain	0.5	1	2
7	99	Painful purulent conjunctivitis	1	1	2
8	40	Severe breathing problems	0.3	1	2

Figure 6: Table of tularemia's symptoms in human.

No	Id	Symptom	Weight	Group	Stat
1	11	Abortion followed by a mucoid, gray green vaginal discharge	1.2	1	0
2	15	Reduced fertility	0.4	1	0
3	16	Weakness in young	0.4	1	0
4	27	Absence of appetite	0.3	1	0
5	101	Inflammation of several joints	0.7	1	0
6	19	Lameness	0.3	1	0
7	11	Abortion	0.8	1	0

Figure 7: Table of brucellosis' symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	97	Ulceration	0.7	1	0
2	27	Absence of appetite	0.4	1	0
3	90	Nasal discharge	0.3	1	0
4	4	Shortness of breath	0.3	1	0
5	2	Fever	0.2	1	0
6	103	Pustules (small inflamed elevation of skin containing pus)	0.7	1	0
7	104	Stellate scar	1	1	0

Figure 8: Table of glanders' symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	42	Flu-like signs	0.2	1	2
2	39	Diarrhea	0.3	1	2
3	60	Chest pain	0.4	1	2
4	103	Pustules (small inflamed elevation of skin containing pus)	0.9	1	2
5	97	Ulceration	0.8	1	2
6	105	Photo-phobia	0.4	1	2
7	91	Enlargement of lymph nodes	0.5	1	2
8	27	Absence of appetite	0.3	1	2
9	4	Shortness of breath	0.3	1	2

Figure 9: Table of glanders' symptoms in human.

No	Id	Symptom	Weight	Group	Stat
1	2	Fever	0.2	1	0
2	27	Absence of appetite	0.2	1	0
3	4	Shortness of breath	0.2	1	0
4	90	Nasal discharge	0.2	1	0
5	31	Mild cough	0.2	1	0
6	91	Enlargement of lymph nodes	0.6	2	0

Figure 10: Table of melioidosis' symptoms in cattle.

No	Id	Symptom	Weight	Group	Stat
1	106	Dermal abscesses	0.6	1	0
2	19	Lameness	0.3	1	0
3	108	Leg swelling	0.4	1	0
4	107	Inflammation of epididymis	0.3	1	0
5	91	Enlargement of lymph nodes	0.6	2	0

Figure 11: Table of melioidosis' symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	42	Flu-like signs	0.2	1	2
2	109	Prostatic abscesses	0.4	1	2
3	60	Chest pain	0.4	1	2
4	103	Pustules (small inflamed elevation of skin containing pus)	0.6	1	2
5	58	Pneumonia	0.4	1	2
6	110	Bronchitis	0.4	1	2
7	91	Enlargement of lymph nodes	0.6	2	2

Figure 12: Table of melioidosis' symptoms in human.

No	Id	Symptom	Weight	Group	Stat
1	4	Shortness of breath	0.3	1	0
2	64	Recumbency(lying down, unable to rise)	0.3	1	0
3	65	Hyperesthesia (increased sensitivity to stimulation, particularly to touch)	0.4	1	0
4	66	Convulsion (seizure)	0.4	1	0
5	67	Paddling (walk unsteadily)	0.3	1	0
6	68	Opisthotonus (a spasm where the head and tail are bend upward and the abdomen is bowed downward)	0.4	1	0
7	69	Loss of consciousness	0.4	1	0
8	70	Bloody diarrhea	1.0	1	0
9	73	Violent purgation	1.5	1	0
10	43	Weakness	0.2	1	0
11	11	Abortion	0.2	1	0
12	12	Milk production drop severely	0.2	1	0

Figure 13: Table of toxins' symptoms in cattle.

No	Id	Symptom	Weight	Group	Stat
1	73	Violent purgation	0.8	1	0
2	70	Bloody diarrhea	0.7	1	0
3	27	Absence of appetite	0.3	1	0
4	43	Weakness	0.3	1	0
5	74	Dehydration	0.4	1	0
6	28	Salivation	0.4	1	0
7	3	Shivering, trembling	0.2	1	0
8	56	Incoordination	0.3	1	0

Figure 14: Table of toxins' symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	42	Flu-like signs	0.2	1	2
2	46	Nausea	0.4	1	2
3	47	Vomiting	0.5	1	2
4	70	Bloody diarrhea	0.8	1	2
5	71	Abdominal cramps	0.3	1	2
6	72	Kidney failure	0.9	1	2
7	41	Shock	0.5	1	2

Figure 15: Table of toxins' symptoms in human.

No	Id	Symptom	Weight	Group	Stat
1	2	Fever	0.3	1	0
2	4	Shortness of breath	0.3	1	0
3	89	Ocular discharge	0.8	1	0
4	90	Nasal discharge	0.8	1	0

Figure 16: Table of nipah's symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	42	Flu-like signs	0.2	1	2
2	2	Fever	0.2	1	2
3	77	Headache	0.2	1	2
4	10	Disorientation	0.3	1	2
5	66	Convulsion (seizure)	0.4	1	2
6	44	Dizziness	0.3	1	2
7	4	Shortness of breath	0.8	1	2
8	53	Muscular pain	0.4	1	2
9	47	Vomiting	0.3	1	2

Figure 17: Table of nipah's symptoms in human.

No	Id	Symptom	Weight	Group	Stat
1	2	Fever	0.2	1	0
2	56	Incoordination	0.3	1	0
3	79	Partial paralysis	0.5	1	0
4	51	Weakness	0.2	1	0

Figure 18: Table of west nile fever's symptoms in cattle.

No	Id	Symptom	Weight	Group	Stat
1	2	Fever	0.2	1	0
2	26	Depression	0.2	1	0
3	66	Convulsion (seizure)	0.5	1	0
4	52	Paralysis	0.7	1	0
5	43	Weakness	0.3	1	0

Figure 19: Table of west nile fever's symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	42	Flu-like signs	0.3	1	2
2	76	Eye pain	0.7	1	2
3	47	Vomiting	0.3	1	2
4	2	Fever	0.4	1	2
5	77	Headache	0.3	1	2
6	78	Neck stiffness	0.6	1	2
7	9	Stupor	0.3	1	2
8	10	Disorientation	0.4	1	2
9	3	Shivering, trembling	0.3	1	2
10	66	Convulsion (seizure)	0.6	1	2

Figure 20: Table of west nile fever's symptoms in human.

No	Id	Symptom	Weight	Group	Stat
1	2	Fever	0.3	1	0
2	27	Absence of appetite	0.4	1	0
3	26	Depression	0.4	1	0
4	43	Weakness	0.3	1	0
5	28	Salivation	0.4	1	0
6	88	Fetid diarrhea	0.8	1	0
7	12	Milk production drop severely	0.3	1	0
8	111	Yellow skin in the newborn	1	1	0
9	112	Abortion storms	1	1	0
10	85	Found dead (in young age)	0.8	1	1
11	84	Hepatic necrosis	1	1	1
12	13	Singeing of the hair (by a history of electrical storm)	0.2	2	1
13	14	Crepitating swelling	0.2	2	1
14	6	Bloody discharge from nose, mouth and anus	0.2	2	1

Figure 21: Table of rift valley fever's symptoms in cattle.

No	Id	Symptom	Weight	Group	Stat
1	2	Fever	0.3	1	0
2	112	Abortion storms	0.7	1	0
3	85	Found dead (in young age)	0.9	1	1
4	84	Hepatic necrosis	1	1	1
5	13	Singeing of the hair (by a history of electrical storm)	0.2	2	1
6	6	Bloody discharge from nose, mouth and anus	0.2	2	1

Figure 22: Table of rift valley fever's symptoms in dog.

No	Id	Symptom	Weight	Group	Stat
1	42	Flu-like signs	0.2	1	2
2	2	Fever	0.3	1	2
3	77	Headache	0.2	1	2
4	43	Weakness	0.3	1	2
5	44	Dizziness	0.2	1	2
6	75	Weight loss	0.4	1	2
7	86	Petechiae (a small purplish spot on a body surface)	0.8	1	2
8	87	Retinopathy	0.7	1	2
9	53	Muscular pain	0.4	1	2
10	78	Neck stiffness	0.3	1	2
11	105	Photo-phobia	0.3	1	2
12	47	Vomiting	0.3	1	2

Figure 23: Table of rift valley fever's symptoms in human.

APPENDIX B

Appendix B contains the red flag value of our system using the information about highly-threatening diseases from Chapter 4 (Figures 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, and 4.8) and Appendix A.

Query 1.

Species: Cattle

Symptoms when the animal is alive:

- Shivering, trembling
- Weakness
- Violent purgation
- Bloody diarrhea

Results:

No	Disease	Red Flag Value
1	Anthrax	0.10344827586206896
2	Botulism	0.16666666666666666
3	Plague	0.0
4	Tularemia	0.0
5	Brucellosis	0.0
6	Glanders	0.0
7	Melioidosis	0.0
8	Q-Fever	0.0
9	Toxins	0.48214285714285715
10	Nipah	0.0
11	West Nile Fever	0.0
12	Rift Valley Fever	0.061224489795918367

Query 2.

Species: Dog

Symptoms when the animal is alive:

- Abnormally rapid heart rate
- Very rapid respiration

- Stiffness

Results:

No	Disease	Red Flag Value
1	Anthrax	0.0
2	Botulism	0.0
3	Plague	0.0
4	Tularemia	0.6666666666666667
5	Brucellosis	0.0
6	Glanders	0.0
7	Melioidosis	0.0
8	Q-Fever	0.0
9	Toxins	0.0
10	Nipah	0.0
11	West Nile Fever	0.0
12	Rift Valley Fever	0.0

Query 3.

Species: Cattle

Symptoms when the animal is alive:

- Fever
- Absence of appetite
- Depression
- Yellow skin in the newborn
- Fetid diarrhea

Results:

No	Disease	Red Flag Value
1	Anthrax	0.06896551724137931
2	Botulism	-0.7333333333333334
3	Plague	0.0
4	Tularemia	0.0
5	Brucellosis	0.0
6	Glanders	0.0
7	Melioidosis	0.4
8	Q-Fever	0.0
9	Toxins	0.0
10	Nipah	0.0
11	West Nile Fever	0.16666666666666669
12	Rift Valley Fever	0.5918367346938777

Query 4.

Species: Dog

Symptoms when the animal is alive:

- Fever
- Shivering, trembling
- Shortness of breath

Results:

No	Disease	Red Flag Value
1	Anthrax	0.0
2	Botulism	0.261904761904762
3	Plague	0.3333333333333333
4	Tularemia	0.060606060606060615
5	Brucellosis	0.0
6	Glanders	0.1388888888888889
7	Melioidosis	0.0
8	Q-Fever	0.0
9	Toxins	0.05882352941176471
10	Nipah	0.2727272727272727
11	West Nile Fever	0.10526315789473685
12	Rift Valley Fever	0.3

Query 5.

Species: Dog

Symptoms when the animal is alive:

- Absence of appetite
- Difficulty swallowing
- Weakness

Results:

No	Disease	Red Flag Value
1	Anthrax	0.0
2	Botulism	0.38095238095238093
3	Plague	0.0
4	Tularemia	0.12121212121212123
5	Brucellosis	0.07317073170731708
6	Glanders	0.1111111111111111
7	Melioidosis	0.0
8	Q-Fever	0.0
9	Toxins	0.17647058823529413
10	Nipah	0.0
11	West Nile Fever	0.15789473684210525
12	Rift Valley Fever	0.0

Query 6.

Species: Cattle

Symptoms when the animal is alive:

- Abortion
- Hygromatous swellings especially of the knees (looks like a water-bag)
- Retained placenta

Results:

No	Disease	Red Flag Value
1	Anthrax	0.06896551724137931
2	Botulism	0.0
3	Plague	0.0
4	Tularemia	0.0
5	Brucellosis	0.608695652173913
6	Glanders	0.0
7	Melioidosis	0.0
8	Q-Fever	0.2631578947368421
9	Toxins	0.03571428571428571
10	Nipah	0.0
11	West Nile Fever	0.0
12	Rift Valley Fever	0.0

ACKNOWLEDGEMENTS

This work would not have been possible without the support of many people. First, I would like to thank my major professor Dr. Les Miller for his guidance, support, and patience while working on this project. I would also like to thank Dr. Gadia and Dr. Roth for their willingness to serve on my POS committee and for their help at the time I need it. Last but certainly not least, my deepest thanks go to my parents, Eddy and Dian, for their continuous love and support.

Melissa Yahya